

09/884948

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

CM protein - protein search, using SW model

Run on: September 22, 2003, 12:39:29 / Search time 44 Seconds
(without alignments)
1194.057 Million cell updates/sec

Title: US-09-884-948-1
Perfect score: 1811
Sequence: 1 DSDRYTPPAEPLDMPDPY.....IFIPKSNTPADPKVKGMP 331

Scoring table: BLOSUM62
Gapop 10.0, Gapext 0.5

Searched: 1107863 seqs, 158726573 residues
Total number of hits satisfying chosen parameters: 1107863

Minimum DB seq length: 0
Maximum DB seq length: 200000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database: A_Geneseq_19Jun03.*
1: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.*
2: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.*
3: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.*
4: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.*
5: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.*
6: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.*
7: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.*
8: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.*
9: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.*
10: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.*
11: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.*
12: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.*
13: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.*
14: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.*
15: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.*
16: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.*
17: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.*
18: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.*
19: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.*
20: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.*
21: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.*
22: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.*
23: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.*
24: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA2003.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

result No.	Score	Query Match Length	ID	Description
1	1811	100.0	331	20 AAY33662
2	1811	100.0	331	20 AAY33665
3	1811	100.0	331	20 AAW67770
4	1811	100.0	331	22 AAB81161
5	1811	100.0	331	23 AAB06742
6	1811	100.0	332	20 AAW67771
7	1811	100.0	332	21 AAB12809
8	1811	100.0	332	23 AAB06743
9	1811	100.0	346	15 AAR43048

10	1811	100.0	406	13 AAR22651	Transglutaminase.
11	1811	100.0	407	22 AAB97831	S. mobaraense IPO
12	1811	100.0	407	22 AAB81166	Prepro-transglutamin
13	1811	100.0	407	24 ABO07390	Foreign protein #1
14	1811	100.0	408	15 AAR49049	BTG-contg. sequenc
15	1802	99.5	331	13 AAR22653	Transglutaminase (
16	1493	82.4	400	22 AAB47007	Transglutaminase.
17	1487	82.1	330	22 AAB81164	Transglutaminase x
18	1487	82.1	416	22 AAB97830	S. cinamomeum IPO
19	1487	82.1	416	24 ABO07393	Foreign protein #3
20	1465	80.9	331	17 AAR81570	Transglutaminase.
21	1465	80.9	331	23 AAB06745	Streptomyces lydic
22	1432	79.1	330	23 AAB06744	Streptovorticillium
23	1391	76.8	328	20 AAY33666	Streptovorticillium
24	162	8.9	39	17 AAR88162	Transglutaminase N
25	140	7.7	30	17 AAR88199	peptide fragment o
26	130	7.2	26	17 AAR88174	peptide fragment o
27	124	6.8	25	17 AAR88175	peptide fragment o
28	119	6.6	23	17 AAR88200	peptide fragment o
29	116	6.4	23	17 AAR88176	peptide fragment o
30	114.5	6.3	889	22 ABB65872	Drosophila melanog
31	114.5	6.3	1217	24 ABB60377	Bombyx mori serici
32	112.5	6.2	1161	22 AAB70419	Drosophila melanog
33	110	6.1	736	22 AAB59845	AAV3B capsid prote
34	109	6.0	21	17 AAR88202	peptide fragment o
35	108	6.0	558	23 AAB05719	Human testis deriv
36	108	6.0	1005	21 AAB43890	Human cancer assoc
37	108	6.0	1182	18 AAW11667	Human metastasis-a
38	107	5.9	736	22 AAB59846	AAV3B capsid prote
39	106.5	5.9	736	21 AAW11667	Adeno-associated v
40	106.5	5.9	736	22 AAB59847	AAV6 capsid protei
41	106	5.8	23	17 AAR88188	peptide fragment o
42	105	5.8	764	21 AAG53024	Arabidopsis thalia
43	105	5.8	906	21 AAG53023	Arabidopsis thalia
44	105	5.8	919	21 AAG53022	Arabidopsis thalia
45	104	5.7	24	17 AAR88169	peptide fragment o

ALIGNMENTS

RESULT 1
AAY33662
ID AAY33662 standard; Protein; 331 AA.
AC AAY33662;
XX
XX 07-JAN-2000 (first entry)
XX
XX Streptovorticillium S-8112 transglutaminase protein fragment.
DE
XX Streptovorticillium
XX Transglutaminase; food industry; pharmaceutical industry; texture;
KW cosmetic industry; proteinaceous material; gel strength; viscosity;
KW breaking strength; elasticity; taste.
XX
XX Streptovorticillium sp.
XX
XX DE19814860-A1.
XX
XX 07-OCT-1999.
XX
XX 02-APR-1998; 98DE-1014860.
XX
XX 02-APR-1998; 98DE-1014860.
XX
XX 02-APR-1998; 98DE-1014860.
XX
XX (FUCH) FUCHSBAUER H.
XX Fuchsbaue H. Pasternack R. Dorsch S. Otterbach J. Roberek I.
XX Matusch M. Dauscher C/
XX
XX WPI; 1999-552288/47.
XX
XX N-PSDB; AAZ23653.
XX

PT Bacterial transglutaminase polypeptides useful for polymerizing
PT proteins, e.g. to modify the properties of food, pharmaceutical or
XX cosmetic products -
XX
PS Claim 1; Page 23-24; 44pp; German.
XX
CC This invention describes a novel bacterial transglutaminase polypeptide.
CC (I) can be used, e.g. in the food, pharmaceutical and cosmetic
CC industries, to polymerize proteinaceous materials in order to improve
CC their properties, e.g. texture, gel strength, breaking strength,
CC viscosity, elasticity or taste. (I) can also be used to immobilize
CC enzymes and antibodies. This sequence represents a transglutaminase
CC isolated from Streptococcus thermophilus S-8112.
XX
SQ Sequence 331 AA;
Query Match 100.0%; Score 1811; DB 20; Length 331;
Best Local Similarity 100.0%; Pred. No. 2.9e-153;
Matches 331; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 DSDRVTTPAEPPLDMPDPYRPSYGRATVNNYIRKQOVSHEHGKQOMTEBQREWL 60
DB 1 DSDRVTTPAEPPLDMPDPYRPSYGRATVNNYIRKQOVSHEHGKQOMTEBQREWL 60
QY 61 SYGCVGVTWNSGQVPTNRLAFASFDREPKNELKGRPRSGETPAEFGVAKESFDEE 120
DB 61 SYGCVGVTWNSGQVPTNRLAFASFDREPKNELKGRPRSGETPAEFGVAKESFDEE 120
QY 121 KGFQRAEVASVMMRLAENAHDESAYLDNLKELANGNDALRNEDARSPFYSALRNTPSF 180
DB 121 KGFQRAEVASVMMRLAENAHDESAYLDNLKELANGNDALRNEDARSPFYSALRNTPSF 180
QY 181 KERNGNHDPSSRMKAVIYSKHFWSGQDRSSADKRYGDPDAFRPAPGTGLVMSRDRI 240
DB 181 KERNGNHDPSSRMKAVIYSKHFWSGQDRSSADKRYGDPDAFRPAPGTGLVMSRDRI 240
QY 241 PRSPTSGEGFVNPDYGFAGQTEADADKVTWTHGNHYAPNGSLGAHAYVESKFRWSE 300
DB 241 PRSPTSGEGFVNPDYGFAGQTEADADKVTWTHGNHYAPNGSLGAHAYVESKFRWSE 300
QY 301 GYSDPDRGAVITTFIPKSWNTAPDKVKQGW 331
DB 301 GYSDPDRGAVITTFIPKSWNTAPDKVKQGW 331
RESULT 2
AAV33665
ID AAV33665 standard; Protein; 331 AA.
XX
AC AAV33665;
DT 07-JAN-2000 (first entry)
XX
DE Streptococcus thermophilus mobaraense transglutaminase protein fragment.
XX
KM Transglutaminase; food industry; pharmaceutical industry; texture;
KM cosmetic industry; proteinaceous material; gel strength; viscosity;
XX breaking strength; elasticity; taste.
XX
OS Streptococcus thermophilus mobaraense.
XX
PN DE19814860-A1.
XX
PD 07-OCT-1999.
XX
PF 02-APR-1998; 98DE-1014860.
XX
PR 02-APR-1998; 98DE-1014860.
XX (FUCHS) FUCHSBAUER H.
XX
PA Fuchsbauer H., Pasternack R., Dorsch S., Otterbach J., Robenek I;
PI Mainusch M., Dauscher C;

XX
DR WP1; 1999-552288/47.
XX
PT Bacterial transglutaminase polypeptides useful for polymerizing
PT proteins, e.g. to modify the properties of food, pharmaceutical or
XX cosmetic products -
XX
PS Claim 12; Page 25-26; 44pp; German.
XX
CC This invention describes a novel bacterial transglutaminase polypeptide.
CC (I) can be used, e.g. in the food, pharmaceutical and cosmetic
CC industries, to polymerize proteinaceous materials in order to improve
CC their properties, e.g. texture, gel strength, breaking strength,
CC viscosity, elasticity or taste. (I) can also be used to immobilize
CC enzymes and antibodies. This sequence represents a transglutaminase
CC protein fragment isolated from Streptococcus thermophilus mobaraense.
XX
SQ Sequence 331 AA;
Query Match 100.0%; Score 1811; DB 20; Length 331;
Best Local Similarity 100.0%; Pred. No. 2.9e-153;
Matches 331; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 DSDRVTTPAEPPLDMPDPYRPSYGRATVNNYIRKQOVSHEHGKQOMTEBQREWL 60
DB 1 DSDRVTTPAEPPLDMPDPYRPSYGRATVNNYIRKQOVSHEHGKQOMTEBQREWL 60
QY 61 SYGCVGVTWNSGQVPTNRLAFASFDREPKNELKGRPRSGETPAEFGVAKESFDEE 120
DB 61 SYGCVGVTWNSGQVPTNRLAFASFDREPKNELKGRPRSGETPAEFGVAKESFDEE 120
QY 121 KGFQRAEVASVMMRLAENAHDESAYLDNLKELANGNDALRNEDARSPFYSALRNTPSF 180
DB 121 KGFQRAEVASVMMRLAENAHDESAYLDNLKELANGNDALRNEDARSPFYSALRNTPSF 180
QY 181 KERNGNHDPSSRMKAVIYSKHFWSGQDRSSADKRYGDPDAFRPAPGTGLVMSRDRI 240
DB 181 KERNGNHDPSSRMKAVIYSKHFWSGQDRSSADKRYGDPDAFRPAPGTGLVMSRDRI 240
QY 241 PRSPTSGEGFVNPDYGFAGQTEADADKVTWTHGNHYAPNGSLGAHAYVESKFRWSE 300
DB 241 PRSPTSGEGFVNPDYGFAGQTEADADKVTWTHGNHYAPNGSLGAHAYVESKFRWSE 300
QY 301 GYSDPDRGAVITTFIPKSWNTAPDKVKQGW 331
DB 301 GYSDPDRGAVITTFIPKSWNTAPDKVKQGW 331
RESULT 3
AAW67770
ID AAW67770 standard; Protein; 331 AA.
XX
AC AAW67770;
DT 01-APR-1999 (first entry)
XX
DE A transglutaminase enzyme sequence.
XX
KM Transglutaminase; microbial; gelled food; jelly; yogurt; cheese;
KM cosmetic; meat quality; microcapsule production; high thermal stability;
XX carrier; immobilised enzyme.
XX
OS Streptococcus thermophilus sp.
XX
PN EP889133-A2.
XX
PD 07-JAN-1999.
XX
PF 02-JUL-1998; 98EP-0112315.
XX
PR 04-JUL-1997; 97JP-0180010.
XX (AJIN) AJINOMOTO CO INC.
PA

XX Miwa T, Nakamura N, Seguro K, Yokoyama K;
 XX WPI; 1999-062664/06.
 DR N-PSDB; AAV81507.
 XX
 XX New microbial transglutaminase with N-terminal aspartic acid deleted
 PT - allowing high level recombinant production without added
 PT methionine in *E. coli*, useful in production of gelled foods,
 PT cosmetics etc.

PS Claim 1; Page 12-14; 56pp; English.

CC The present sequence represents a transglutaminase of *Streptococcus*
 CC *sp.*. The specification describes a new microbial transglutaminase that
 CC has the N-terminal aspartic acid of the present transglutaminase
 CC deleted. Eliminating the N-terminal Asp from microbial transglutaminase
 CC allows efficient removal of the terminal methionine residue added when
 CC the protein is expressed in *Escherichia coli*. The *E. coli* methionine
 CC aminopeptidase acts well on Met-Ser but only poorly on Met-Asp, so
 CC problems of antigenicity associated with Met-terminated proteins are
 CC avoided. Recombinant transglutaminase is used to produce gelled foods
 CC (jellies, yogurt and cheeses) or cosmetics, to improve the quality of
 CC meat, in the production of materials for microcapsules of high thermal
 CC stability and as a carrier for immobilised enzymes.

XX Sequence 331 AA;

Query Match 100.0%; Score 1811; DB 20; Length 331;
 Best Local Similarity 100.0%; Pred. No. 2.9e-153;
 Matches 331; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DSDRTPPAPPELDMPDPYRPSYGRATVNNYIRKQOVYSHRDKQOMTEOREWL 60
 DB 1 DSDRTPPAPPELDMPDPYRPSYGRATVNNYIRKQOVYSHRDKQOMTEOREWL 60
 QY 61 SYGCVGVTVWNSGQYPTNRLAFASFDDEDFKNELKNGRPSGETRAEFGRVAKESFDEE 120
 DB 61 SYGCVGVTVWNSGQYPTNRLAFASFDDEDFKNELKNGRPSGETRAEFGRVAKESFDEE 120
 QY 121 KGPQRAREVASVNNRLENAHDESAVLDNLKELANGNDALNENARSPYSALRNTPSF 180
 DB 121 KGPQRAREVASVNNRLENAHDESAVLDNLKELANGNDALNENARSPYSALRNTPSF 180
 QY 181 KERNGNHDPSSRMKAVIYSKHFWSGODRSSADKRYGDPDAFRPAPGTGLVDMGRDNT 240
 DB 181 KERNGNHDPSSRMKAVIYSKHFWSGODRSSADKRYGDPDAFRPAPGTGLVDMGRDNT 240
 QY 241 PRSPTPSGEGFVNFYGFAGQTEADADKTVTHGNHHAHNGSLGAMHYESKFRMWS 300
 DB 241 PRSPTPSGEGFVNFYGFAGQTEADADKTVTHGNHHAHNGSLGAMHYESKFRMWS 300
 QY 301 GYSDPDRGAYVITFIKSNMTAPDKVKGMP 331
 DB 301 GYSDPDRGAYVITFIKSNMTAPDKVKGMP 331

RESULT 4

AA881161
 ID AA881161 standard; protein; 331 AA.

XX AA881161;

DT 13-JUN-2001 (first entry)

DE Transglutaminase related protein SEQ ID 5.

KM Coryneform bacteria; transglutaminase; food processing.

XX Streptococcus mobaraense.

XX WO200123591-A1.

XX

PD 05-APR-2001.

PF 29-SEP-2000; 2000WO-JP06780.

PR 30-SEP-1999; 99JP-0280098.

PR 28-JUN-2000; 2000JP-0194043.

XX (AJIN) AJINOMOTO CO INC.

PI Kikuchi Y, Date M, Umezawa Y, Yokoyama K, Matsui H;

DR WPI; 2001-266172/27.

PT Efficient secretory production of foreign proteins e.g.
 PT transglutaminase employing transformant coryneform bacterium, simply on
 PT industrial scale with direct recovery for use in food processing and
 PT pharmaceutical industry

PS Claim 22; Page 79-81; 151pp; Japanese.

CC This invention relates to a process for the production of a foreign
 CC secretory protein through the construction of a recombinant coryneform
 CC bacterium. The coryneform bacterium is transformed with an expression
 CC construct in which DNA encoding a target foreign protein pro-structure is
 CC ligated to the downstream region of DNA encoding the signal peptide
 CC domain of a coryneform bacterial protein. Following transformation with
 CC the vector, the bacterium is cultured, and the pro-peptide cleaved from
 CC the expressed protein. Transglutaminases produced using this process are
 CC useful in the food processing and pharmaceutical industries. The present
 CC sequence represents a transglutaminase related protein, which can be
 CC used in the method of the invention.

XX Sequence 331 AA;

Query Match 100.0%; Score 1811; DB 22; Length 331;
 Best Local Similarity 100.0%; Pred. No. 2.9e-153;
 Matches 331; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DSDRTPPAPPELDMPDPYRPSYGRATVNNYIRKQOVYSHRDKQOMTEOREWL 60
 DB 1 DSDRTPPAPPELDMPDPYRPSYGRATVNNYIRKQOVYSHRDKQOMTEOREWL 60
 QY 61 SYGCVGVTVWNSGQYPTNRLAFASFDDEDFKNELKNGRPSGETRAEFGRVAKESFDEE 120
 DB 61 SYGCVGVTVWNSGQYPTNRLAFASFDDEDFKNELKNGRPSGETRAEFGRVAKESFDEE 120
 QY 121 KGPQRAREVASVNNRLENAHDESAVLDNLKELANGNDALNENARSPYSALRNTPSF 180
 DB 121 KGPQRAREVASVNNRLENAHDESAVLDNLKELANGNDALNENARSPYSALRNTPSF 180
 QY 181 KERNGNHDPSSRMKAVIYSKHFWSGODRSSADKRYGDPDAFRPAPGTGLVDMGRDNT 240
 DB 181 KERNGNHDPSSRMKAVIYSKHFWSGODRSSADKRYGDPDAFRPAPGTGLVDMGRDNT 240
 QY 241 PRSPTPSGEGFVNFYGFAGQTEADADKTVTHGNHHAHNGSLGAMHYESKFRMWS 300
 DB 241 PRSPTPSGEGFVNFYGFAGQTEADADKTVTHGNHHAHNGSLGAMHYESKFRMWS 300
 QY 301 GYSDPDRGAYVITFIKSNMTAPDKVKGMP 331
 DB 301 GYSDPDRGAYVITFIKSNMTAPDKVKGMP 331

RESULT 5

ABB06742
 ID ABB06742 standard; Protein; 331 AA.

XX ABB06742;

DT 12-JUN-2002 (first entry)

DE Streptococcus mobaraense transglutaminase protein SEQ ID NO:2.

XX

XX	Microbial; transglutaminase; protein co-ordinate data; stereo-structure;
KM	X-ray analysis; crystalline structure; enzyme; food processing;
KW	thermal stability.
OS	Streptomyces mobaraensis.
XX	
XX	MO20021518-AL.
XX	
PD	21-FEB-2002.
XX	
XX	15-AUG-2001; 2001WG-JB07038.
XX	
XX	17-AUG-2000; 2000JP-0247664.
PR	27-DEC-2000; 2000JP-0396695.
XX	
PA	(AJIN) AJINOMOTO CO INC.
PI	Kashiwagi T, Shimba N, Ishikawa K, Suzuki E, Yokoyama K;
PI	Hirayama K;
XX	
XX	WPI: 2002-269198/31.
DR	N-PSDB; ABL50235.
PT	Designing and constructing a variant of Streptococcicillium
PT	mobaraense-originated transglutaminase (MTG), for use in food
PT	processing, comprises estimating the binding site of MTG based
PT	stereo-structure -
XX	
PS	Claim 2; Page 100-102; 126pp; Japanese.
XX	
CC	The present invention describes a method for designing and constructing
CC	a variant transglutaminase by estimating the binding site of
CC	Streptococcicillium mobaraense-originated transglutaminase (MTG) to the
CC	substrate based on the stereo-structure obtained by X-ray analysis of
CC	the crystalline structure of MTG crystals, and e.g. substituting amino
CC	acid residues located at the substrate-binding site of the
CC	transglutaminase. The method can be used for designing and constructing
CC	a variant transglutaminase. The obtained transglutaminases can be used
CC	in food processing. The modified transglutaminases have improved
CC	transglutaminase activity and thermal stability, substrate-specificity
CC	and an less required optimum pH. The present sequence represents a
CC	transglutaminase isolated from Streptococcicillium mobaraense
CC	(also called Streptomyces mobaraensis).
XX	
SQ	Sequence 331 AA;
XX	
Query Match	100.0%; Score 1811; DB 23; Length 331;
Best Local Similarity	100.0%; Pred. No.2,9e-153;
Matches 331; Conservative	0; Mismatches 0; Indels 0; Gaps 0;
QY	1 DSDDRVPPAPLDRPDPYRSGRATTVNNYIRKQOVS SHDGRKQOMTBEORENL 60
DB	1 DSDDRVPPAPLDRMPDPRPSGRATTVNNYIRKQOVS SHDGRKQOMTBEORENL 60
QY	61 SYGCGVTVNNSGOVYPTNLAPASFDEGRFKNELKNGRPSGSETPAFRGRVAKESFDEE 120
DB	61 SYGCGVTVNNSGOVYPTNLAPASFDEGRFKNELKNGRPSGSETPAFRGRVAKESFDEE 120
QY	121 KQFORAREVAVSNMRALENAAHDESAVIDNLKKELANGDNLNEDARSPPYSALRNTPSF 180
DB	121 KQFORAREVAVSNMRALENAAHDESAVIDNLKKELANGDNLNEDARSPPYSALRNTPSF 180
QY	181 KRRNGGNHDEPRKMAZYVYKGFHWSSQDSSADPRKKGPPDAPRAPOTGYVDSSEDRNY 240
DB	181 KRRNGGNHDEPRKMAZYVYKGFHWSSQDSSADPRKKGPPDAPRAPOTGYVDSSEDRNY 240
QY	241 PRSPFSPSEGGVFDYDGMGAQTEADAKTWTGNNHYAPNGSGAHHVYESKFRNWS 300
DB	241 PRSPFSPSEGGVFDYDGMGAQTEADAKTWTGNNHYAPNGSGAHHVYESKFRNWS 300
QY	301 GYSDFPRAGAVYITFIKSNMTAPDKYKGNP 331
DB	301 GYSDFPRAGAVYITFIKSNMTAPDKYKGNP 331

```

ID      RESULT 6
AA067771
AA067771 standard; Protein; 332 AA.
AC      AA067771;
AD      01-APR-1999 (first entry)
AE      XX
AF      XX
AG      XX
AH      XX
AI      XX
AJ      XX
AK      XX
AL      XX
AM      XX
AN      XX
AO      XX
AP      XX
AQ      XX
AR      XX
AS      XX
AT      XX
AU      XX
AV      XX
AW      XX
AX      XX
AY      XX
AZ      XX
BA      XX
BB      XX
BC      XX
BD      XX
BE      XX
BF      XX
BG      XX
BH      XX
BI      XX
BJ      XX
BK      XX
BL      XX
BM      XX
BN      XX
BO      XX
BP      XX
BQ      XX
BR      XX
BS      XX
BT      XX
BU      XX
BV      XX
BW      XX
BX      XX
BY      XX
BZ      XX
CA      XX
CB      XX
CC      XX
CD      XX
CE      XX
CF      XX
CG      XX
CH      XX
CI      XX
CJ      XX
CK      XX
CL      XX
CM      XX
CN      XX
CO      XX
CP      XX
CQ      XX
CR      XX
CS      XX
CT      XX
CU      XX
CV      XX
CW      XX
CX      XX
CY      XX
CZ      XX
DA      XX
DB      XX
DC      XX
DD      XX
DE      XX
DF      XX
DG      XX
DH      XX
DI      XX
DJ      XX
DK      XX
DL      XX
DM      XX
DN      XX
DO      XX
DP      XX
DQ      XX
DR      XX
DS      XX
DT      XX
DU      XX
DV      XX
DW      XX
DX      XX
DY      XX
EZ      XX
FA      XX
FB      XX
FC      XX
FD      XX
FE      XX
FF      XX
FG      XX
FH      XX
FI      XX
FJ      XX
FK      XX
FL      XX
FM      XX
FN      XX
FO      XX
FP      XX
FQ      XX
FR      XX
FS      XX
FT      XX
FU      XX
FV      XX
FW      XX
FX      XX
FY      XX
FZ      XX
GA      XX
GB      XX
GC      XX
GD      XX
GE      XX
GF      XX
GG      XX
GH      XX
GI      XX
GJ      XX
GK      XX
GL      XX
GM      XX
GN      XX
GO      XX
GP      XX
GQ      XX
GR      XX
GS      XX
GT      XX
GU      XX
GV      XX
GW      XX
GX      XX
GY      XX
GZ      XX
HA      XX
HB      XX
HC      XX
HD      XX
HE      XX
HF      XX
HG      XX
HH      XX
HI      XX
HJ      XX
HK      XX
HL      XX
HM      XX
HN      XX
HO      XX
HP      XX
HQ      XX
HR      XX
HS      XX
HT      XX
HU      XX
HV      XX
HW      XX
HX      XX
HY      XX
HZ      XX
IA      XX
IB      XX
IC      XX
ID      XX
IE      XX
IF      XX
IG      XX
IH      XX
II      XX
IJ      XX
IK      XX
IL      XX
IM      XX
IN      XX
IO      XX
IP      XX
IQ      XX
IR      XX
IS      XX
IT      XX
IU      XX
IV      XX
IW      XX
IX      XX
IY      XX
IZ      XX
JA      XX
JB      XX
JC      XX
JD      XX
JE      XX
JF      XX
JG      XX
JH      XX
JI      XX
JJ      XX
JK      XX
JL      XX
JM      XX
JN      XX
JO      XX
JP      XX
JQ      XX
JR      XX
JS      XX
JT      XX
JU      XX
JV      XX
JW      XX
JX      XX
JY      XX
JZ      XX
KA      XX
KB      XX
KC      XX
KD      XX
KE      XX
KF      XX
KG      XX
KH      XX
KI      XX
KJ      XX
KK      XX
KL      XX
KM      XX
KN      XX
KO      XX
KP      XX
KQ      XX
KR      XX
KS      XX
KT      XX
KU      XX
KV      XX
KW      XX
KX      XX
KY      XX
KZ      XX
LA      XX
LB      XX
LC      XX
LD      XX
LE      XX
LF      XX
LG      XX
LH      XX
LI      XX
LJ      XX
LK      XX
LL      XX
LM      XX
LN      XX
LO      XX
LP      XX
LQ      XX
LR      XX
LS      XX
LT      XX
LU      XX
LV      XX
LW      XX
LX      XX
LY      XX
LZ      XX
MA      XX
MB      XX
MC      XX
MD      XX
ME      XX
MF      XX
MG      XX
MH      XX
MI      XX
MJ      XX
MK      XX
ML      XX
MM      XX
MN      XX
MO      XX
MP      XX
MQ      XX
MR      XX
MS      XX
MT      XX
MU      XX
MV      XX
MW      XX
MX      XX
MY      XX
MZ      XX
NA      XX
NB      XX
NC      XX
ND      XX
NE      XX
NF      XX
NG      XX
NH      XX
NI      XX
NJ      XX
NK      XX
NL      XX
NM      XX
NO      XX
NP      XX
NQ      XX
NR      XX
NS      XX
NT      XX
NU      XX
NV      XX
NW      XX
NX      XX
NY      XX
NZ      XX
OA      XX
OB      XX
OC      XX
OD      XX
OE      XX
OF      XX
OG      XX
OH      XX
OI      XX
OJ      XX
OK      XX
OL      XX
OM      XX
ON      XX
OO      XX
OP      XX
OQ      XX
OR      XX
OS      XX
OT      XX
OU      XX
OV      XX
OW      XX
OX      XX
OY      XX
OZ      XX
PA      XX
PB      XX
PC      XX
PD      XX
PE      XX
PF      XX
PG      XX
PH      XX
PI      XX
PJ      XX
PK      XX
PL      XX
PM      XX
PN      XX
PO      XX
PP      XX
PQ      XX
PR      XX
PS      XX
PT      XX
PU      XX
PV      XX
PW      XX
PX      XX
PY      XX
PZ      XX
QA      XX
QB      XX
QC      XX
QD      XX
QE      XX
QF      XX
QG      XX
QH      XX
QI      XX
QJ      XX
QK      XX
QL      XX
QM      XX
QN      XX
QO      XX
QP      XX
QQ      XX
QR      XX
QS      XX
QT      XX
QU      XX
QV      XX
QW      XX
QX      XX
QY      XX
QZ      XX
RA      XX
RB      XX
RC      XX
RD      XX
RE      XX
RF      XX
RG      XX
RH      XX
RI      XX
RJ      XX
RK      XX
RL      XX
RM      XX
RN      XX
RO      XX
RP      XX
RQ      XX
RR      XX
RS      XX
RT      XX
RU      XX
RV      XX
RW      XX
RX      XX
RY      XX
RZ      XX
SA      XX
SB      XX
SC      XX
SD      XX
SE      XX
SF      XX
SG      XX
SH      XX
SI      XX
SJ      XX
SK      XX
SL      XX
SM      XX
SN      XX
SO      XX
SP      XX
SQ      XX
SR      XX
SS      XX
ST      XX
SU      XX
SV      XX
SW      XX
SX      XX
SY      XX
SZ      XX
TA      XX
TB      XX
TC      XX
TD      XX
TE      XX
TF      XX
TG      XX
TH      XX
TI      XX
TJ      XX
TK      XX
TL      XX
TM      XX
TN      XX
TO      XX
TP      XX
TQ      XX
TR      XX
TS      XX
TT      XX
TU      XX
TV      XX
TW      XX
TX      XX
TY      XX
TZ      XX
UA      XX
UB      XX
UC      XX
UD      XX
UE      XX
UF      XX
UG      XX
UH      XX
UI      XX
UJ      XX
UK      XX
UL      XX
UM      XX
UN      XX
UO      XX
UP      XX
UQ      XX
UR      XX
US      XX
UT      XX
UU      XX
UV      XX
UW      XX
UX      XX
UY      XX
UZ      XX
VA      XX
VB      XX
VC      XX
VD      XX
VE      XX
VF      XX
VG      XX
VH      XX
VI      XX
VJ      XX
VK      XX
VL      XX
VM      XX
VN      XX
VO      XX
VP      XX
VQ      XX
VR      XX
VS      XX
VT      XX
VU      XX
VV      XX
VW      XX
VX      XX
VY      XX
VZ      XX
WA      XX
WB      XX
WC      XX
WD      XX
WE      XX
WF      XX
WG      XX
WH      XX
WI      XX
WJ      XX
WK      XX
WL      XX
WM      XX
WN      XX
WO      XX
WP      XX
WQ      XX
WR      XX
WS      XX
WT      XX
WU      XX
WV      XX
WW      XX
WX      XX
WY      XX
WZ      XX
XA      XX
XB      XX
XC      XX
XD      XX
XE      XX
XF      XX
XG      XX
XH      XX
XI      XX
XJ      XX
XK      XX
XL      XX
XM      XX
XN      XX
XO      XX
XP      XX
XQ      XX
XR      XX
XS      XX
XT      XX
XU      XX
XV      XX
XW      XX
XX      XX
XY      XX
XZ      XX
YA      XX
YB      XX
YC      XX
YD      XX
YE      XX
YF      XX
YG      XX
YH      XX
YI      XX
YJ      XX
YK      XX
YL      XX
YM      XX
YN      XX
YO      XX
YP      XX
YQ      XX
YR      XX
YS      XX
YT      XX
YU      XX
YV      XX
YW      XX
YX      XX
YY      XX
YZ      XX
ZA      XX
ZB      XX
ZC      XX
ZD      XX
ZE      XX
ZF      XX
ZG      XX
ZH      XX
ZI      XX
ZJ      XX
ZK      XX
ZL      XX
ZM      XX
ZN      XX
ZO      XX
ZP      XX
ZQ      XX
ZR      XX
ZS      XX
ZT      XX
ZU      XX
ZV      XX
ZW      XX
ZX      XX
ZY      XX
ZZ      XX

```

Db 122 KGFQARERVAIVNRLAENAHDESAVLDNLKKELANGNDALNEDAPSPFYSAIRNTSPF 181
 QY 181 KERNGNHDPSSMKAVIYSKHPWSGODRSSADKRYGDPDAFRAPGTLVDMASDRNT 240
 Db 182 KERNGNHDPSSMKAVIYSKHPWSGODRSSADKRYGDPDAFRAPGTLVDMASDRNT 241
 QY 241 PRSPFSGEGFVNPDYGMFGAQTADKTYWTHGNHYHAPNCSLGAMHYESKFRWMS 300
 Db 242 PRSPFSGEGFVNPDYGMFGAQTADKTYWTHGNHYHAPNCSLGAMHYESKFRWMS 301
 QY 301 GYSDFRGAIVITFIPIKSWNTAPDKVKQGW 331
 Db 302 GYSDFRGAIVITFIPIKSWNTAPDKVKQGW 332

RESULT 7
 AAB12809
 ID AAB12809 standard; Protein; 332 AA.
 AC AAB12809;
 XX
 DT 24-NOV-2000 (first entry)
 DE Transglutaminase protein sequence SEQ ID NO:1.
 XX
 KM Transglutaminase; gelled food; jelly; yoghurt; gelled cosmetic;
 KM cheese.
 XX
 OS Unidentified.
 XX
 PN WO20040706-A1.
 XX
 PD 13-JUL-2000.
 XX
 PF 24-DEC-1999; 99WO-JP07250.
 XX
 PR 28-DEC-1998; 98JP-0373131.
 XX
 PA (AJIN) AJINOMOTO CO INC.
 PI Yokoyama K, Ono K, Ejima D;
 DR WPI; 2000-475826/41.
 DR N-PSDB; AAA73025.
 XX
 PT Production of active transglutaminase from denatured enzyme by
 PT two-stage refolding process for industrial production of active enzyme
 PT for use in food production
 XX
 PS Disclosure; Page 48-50; 74pp; Japanese.
 XX
 CC The present invention describes a method for producing active
 CC transglutaminase from denatured enzyme. The method comprises: (i) forming
 CC an intermediate structure of the enzyme having transglutaminase activity
 CC under acidic conditions in an aqueous medium; and (ii) forming a higher-
 CC level structure of the enzyme having transglutaminase activity under
 CC neutral conditions in an aqueous medium. The method can be used for
 CC industrial production of active transglutaminase from denatured material
 CC (such as recombinant transglutaminase) which can be used in the food
 CC industry for the production of gelled foods such as jellies, yoghurts
 CC and cheeses, and for the production of gelled cosmetics. The present
 CC sequence represents a transglutaminase which is used in the
 CC exemplification from the present invention.
 CC
 XX
 SQ Sequence 332 AA;
 QY Query Match 100.0%; Score 1811; DB 21; Length 332;
 Matches 331; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 1 DSDRVTTPAPLDRMPDPYRPSYGRATVNNYIRKQOQVSHDGRKQOMTEOREWL 60
 2 DSDRVTTPAPLDRMPDPYRPSYGRATVNNYIRKQOQVSHDGRKQOMTEOREWL 61

QY 61 SYGCVGTWNSQGYPTNRLAPAFSDREKRLKNGRPSGETFAEREGVANAESFDE 120
 Db 62 SYGCVGTWNSQGYPTNRLAPAFSDREKRLKNGRPSGETFAEREGVANAESFDE 121
 QY 121 KGFQARERVAIVNRLAENAHDESAVLDNLKKELANGNDALNEDAPSPFYSAIRNTSPF 180
 Db 122 KGFQARERVAIVNRLAENAHDESAVLDNLKKELANGNDALNEDAPSPFYSAIRNTSPF 181
 QY 181 KERNGNHDPSSMKAVIYSKHPWSGODRSSADKRYGDPDAFRAPGTLVDMASDRNT 240
 Db 182 KERNGNHDPSSMKAVIYSKHPWSGODRSSADKRYGDPDAFRAPGTLVDMASDRNT 241
 QY 241 PRSPFSGEGFVNPDYGMFGAQTADKTYWTHGNHYHAPNCSLGAMHYESKFRWMS 300
 Db 242 PRSPFSGEGFVNPDYGMFGAQTADKTYWTHGNHYHAPNCSLGAMHYESKFRWMS 301
 QY 301 GYSDFRGAIVITFIPIKSWNTAPDKVKQGW 331
 Db 302 GYSDFRGAIVITFIPIKSWNTAPDKVKQGW 332

RESULT 8
 ABB06743
 ID ABB06743 standard; Protein; 332 AA.
 AC ABB06743;
 XX
 DT 12-JUN-2002 (first entry)
 DE Streptococcus mobaraense transglutaminase protein SEQ ID NO:4.
 XX
 KM Microbial; transglutaminase; protein co-ordinate data; stereo-structure;
 KM X-ray analysis; crystalline structure; enzyme; food processing;
 KM thermal stability.
 XX
 OS Streptococcus mobaraensis.
 XX
 FH Key Location/Qualifiers
 FT Peptide 1
 FT Protein /label= signal
 FT /label= transglutaminase
 PN WO200214518-A1.
 XX
 PD 21-FEB-2002.
 XX
 PF 15-AUG-2001; 2001WO-JP07038.
 XX
 PR 17-AUG-2000; 2000JP-0247664.
 PR 27-DEC-2000; 2000JP-0396695.
 XX
 PA (AJIN) AJINOMOTO CO INC.
 PI Kashiwagi T, Shimba N, Ishikawa K, Suzuki E, Yokoyama K;
 PI Hiiyama K;
 DR WPI; 2002-269198/31.
 DR N-PSDB; ABL50236.
 XX
 PT Designing and constructing a variant of Streptococcus
 PT mobaraense-originated transglutaminase (MTG), for use in food
 PT processing, comprises estimating the binding site of MTG based on its
 PT stereo-structure
 XX
 PS Example 4; Page 105-106; 126pp; Japanese.
 XX
 CC The present invention describes a method for designing and constructing
 CC a variant transglutaminase by estimating the binding site of
 CC Streptococcus mobaraense-originated transglutaminase (MTG) to the
 CC substrate based on the stereo-structure obtained by X-ray analysis of
 CC the crystalline structure of MTG crystals, and e.g. substituting amino

CC acid residues located at the substrate-binding site of the
 CC transglutaminase. The method can be used for designing and constructing
 CC a variant transglutaminase. The obtained transglutaminases can be used
 CC in food processing. The modified transglutaminases have improved
 CC transglutaminase activity and thermal stability, substrate-specificity
 CC and an less required optimum pH. The present sequence represents a
 CC transglutaminase isolated from *Streptococcus thermophilus* mobarensis
 CC (also called *Streptococcus mobarensis*).
 CC
 XX Sequence 332 AA;
 SQ
 Query Match 100.0%; Score 1811; DB 23; Length 332;
 Best Local Similarity 100.0%; Pred. No. 2.9e-153;
 Matches 331; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 DSDDRVTTPPAEPLDRMPDPRPSYGRATVNNYIRKQVYSHDRKQKQWTEQREWL 60
 DB 2 DSDDRVTTPPAEPLDRMPDPRPSYGRATVNNYIRKQVYSHDRKQKQWTEQREWL 61
 QY 61 SYGCVGVTWNSGQYPTNRLAFASFEDEEFKNEELKNGRPSGRTAEFGRAKESFDEE 120
 DB 62 SYGCVGVTWNSGQYPTNRLAFASFEDEEFKNEELKNGRPSGRTAEFGRAKESFDEE 121
 QY 121 KGFQARREVASVNNRRALENVAHDESAYLNLKKELANGNDALRNEDARSFFYSALRNTPSF 180
 DB 122 KGFQARREVASVNNRRALENVAHDESAYLNLKKELANGNDALRNEDARSFFYSALRNTPSF 181
 QY 181 KERNGNHDPERRKAVIYKHFWSGQDRSSADRKXGDPDARPPAPGTGLVDSRDNI 240
 DB 182 KERNGNHDPERRKAVIYKHFWSGQDRSSADRKXGDPDARPPAPGTGLVDSRDNI 241
 QY 241 PRSPFSPGEGFVNFDYGMFGAQTEDADKVTWTHGNHYHAPNGSLGAMVYESKFRMNS 300
 DB 242 PRSPFSPGEGFVNFDYGMFGAQTEDADKVTWTHGNHYHAPNGSLGAMVYESKFRMNS 301
 QY 301 GYSDFDRGAVYITFIPIKSWNTAPDKVKQGW 331
 DB 302 GYSDFDRGAVYITFIPIKSWNTAPDKVKQGW 332

RESULT 9
 AAR49048
 ID AAR49048 standard; Protein; 346 AA.
 XX
 AC AAR49048;
 XX
 DT 10-MAR-2003 (updated)
 DT 20-SEP-1994 (first entry)
 XX
 DE Bacterial transglutaminase.
 DE
 KM Bacterial transglutaminase; BTG; expression;
 KM active; inactive; inclusion body.
 XX
 OS Escherichia coli.
 OS Synthetic.
 XX
 PN JP06030771-A.
 XX
 PD 08-FEB-1994.
 XX
 PF 14-JUL-1992; 92JP-0187038.
 XX
 PR 14-JUL-1992; 92JP-0187038.
 XX
 PA (AJIN) AJINOMOTO KK.
 XX
 DR WPI, 1994-079294/10.
 DR N-PSDB; AAO55983.
 XX
 PT Prepn. of bacterial trans-glutaminase in large ants. - by
 PT expression of fusion protein in E. coli bacterial
 PT trans-glutaminase

XX
 PS Claim 1; Page 7-8; 13pp; Japanese.
 XX
 CC A fused protein contains amino acids 16-346 of BTG (AAO55983)
 CC and a hydrophilic peptide at the amino terminal.
 CC Expression of DNA encoding this protein in E. coli allow large
 CC scale prodn. of BTG. An active BTG can be prepd. from the
 CC inactive fused protein inclusion body.
 CC (Updated on 10-MAR-2003 to add missing OS field.)
 CC
 XX Sequence 346 AA;
 SQ
 Query Match 100.0%; Score 1811; DB 15; Length 346;
 Best Local Similarity 100.0%; Pred. No. 3.1e-153;
 Matches 331; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 DSDDRVTTPPAEPLDRMPDPRPSYGRATVNNYIRKQVYSHDRKQKQWTEQREWL 60
 DB 16 DSDDRVTTPPAEPLDRMPDPRPSYGRATVNNYIRKQVYSHDRKQKQWTEQREWL 75
 QY 61 SYGCVGVTWNSGQYPTNRLAFASFEDEEFKNEELKNGRPSGRTAEFGRAKESFDEE 120
 DB 76 SYGCVGVTWNSGQYPTNRLAFASFEDEEFKNEELKNGRPSGRTAEFGRAKESFDEE 135
 QY 121 KGFQARREVASVNNRRALENVAHDESAYLNLKKELANGNDALRNEDARSFFYSALRNTPSF 180
 DB 136 KGFQARREVASVNNRRALENVAHDESAYLNLKKELANGNDALRNEDARSFFYSALRNTPSF 195
 QY 181 KERNGNHDPERRKAVIYKHFWSGQDRSSADRKXGDPDARPPAPGTGLVDSRDNI 240
 DB 196 KERNGNHDPERRKAVIYKHFWSGQDRSSADRKXGDPDARPPAPGTGLVDSRDNI 255
 QY 241 PRSPFSPGEGFVNFDYGMFGAQTEDADKVTWTHGNHYHAPNGSLGAMVYESKFRMNS 300
 DB 256 PRSPFSPGEGFVNFDYGMFGAQTEDADKVTWTHGNHYHAPNGSLGAMVYESKFRMNS 315
 QY 301 GYSDFDRGAVYITFIPIKSWNTAPDKVKQGW 331
 DB 316 GYSDFDRGAVYITFIPIKSWNTAPDKVKQGW 346

RESULT 10
 AAR22651
 ID AAR22651 standard; Protein; 406 AA.
 XX
 AC AAR22651;
 XX
 DT 25-MAR-2003 (updated)
 DT 09-OCT-1992 (first entry)
 XX
 DE Transglutaminase.
 DE
 KM BTG; acyl rearrangement; deamination.
 KM
 XX Actinomycetes and Streptococcus.
 OS
 OS Key
 FH Peptide
 FT Location/Qualifiers
 FT 1..75
 FT /label= sig_peptide
 FT 76..406
 FT /label= mat_BTG
 FT
 XX EP481504-A.
 XX
 XX 22-APR-1992.
 XX
 PF 18-OCT-1991; 91EP-0117813.
 XX
 PR 19-OCT-1990; 90JP-0282566.
 XX
 PA (AMAN) AMANO PHARM KK.
 XX
 XX (AJIN) AJINOMOTO KK.

PI Ando K, Arafuka S, Koikeda S, Matsui H, Takagi H, Washizu K;
 XX WPI; 1992-133808/17.
 DR N-PSDB; AAQ24197, AAQ24201.
 XX
 PT DNA fragment encoding trans:glutaminase - is inserted into
 PT vector, e.g. Pml1053-BTG, for protein expression
 XX
 PS Disclosure; Page 42 and 44; 55 pp; English.
 CC The mature transglutaminase enzyme (BTG) can be derived from two
 CC different species, Streptococcus lactis and Actinomyces. It
 CC catalyzes an acyl rearrangement reaction of a gamma-carboxamide gp.
 CC of glutamine. It introduces intra- or intermolecular formation of
 CC epsilon-(gamma-Gln)-Lys cross-linking when an epsilon-amino gp. of
 CC a Lys residue acts as an acyl receptor. When water acts as an acyl
 CC acceptor the enzyme accelerates the conversion of Gln residues to Gln
 CC residues by deamination.
 CC The enzyme is used in the prodn. of gelled foods, gelled cosmetics,
 CC yogurt, gelatin, cheese etc. It is also used in the prodn. of
 CC thermally stable materials such as microcapsules and carriers of
 CC immobilized enzymes. The DNA sequence given allows the prodn. of
 CC BTG efficiently and in large quantity.
 CC (Updated on 25-MAR-2003 to correct PA field.)
 CC
 XX SQ Sequence 406 AA;
 Query Match 100.0%; Score 1811; DB 13; Length 406;
 Best Local Similarity 100.0%; Pred. No. 3.8e-153;
 Matches 331; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 DSDDRVTPPAEPLDRMPDPRPSYGRATVNNYIRKQOVYSHRDGRKQMTTEOREWL 60
 DB DSDDRVTPPAEPLDRMPDPRPSYGRATVNNYIRKQOVYSHRDGRKQMTTEOREWL 135
 QY 61 SYCGVGVTVWNSGOYPTNRLAFSPDEDFKNEKXGPRSGETRAEFGVAKESFDEE 120
 DB SYCGVGVTVWNSGOYPTNRLAFSPDEDFKNEKXGPRSGETRAEFGVAKESFDEE 195
 QY 121 KGFORAREVAVNNRBALENHDSAYLDNLKKELANGDALARNDARSFFYSALNTPSF 180
 DB KGFORAREVAVNNRBALENHDSAYLDNLKKELANGDALARNDARSFFYSALNTPSF 255
 QY 196 KGFORAREVAVNNRBALENHDSAYLDNLKKELANGDALARNDARSFFYSALNTPSF 255
 DB KGFORAREVAVNNRBALENHDSAYLDNLKKELANGDALARNDARSFFYSALNTPSF 315
 QY 181 KERNGNNDPSRMKAVIYSKHPWSGQDRSSADKRYGDPDAPRPAOTGLVMSRDNI 240
 DB KERNGNNDPSRMKAVIYSKHPWSGQDRSSADKRYGDPDAPRPAOTGLVMSRDNI 315
 QY 241 PRSPTSPGEGFVNDYGMFGAQTADADKTWTHGNHYAPNGSLGAMHYESKFRNMS 300
 DB PRSPTSPGEGFVNDYGMFGAQTADADKTWTHGNHYAPNGSLGAMHYESKFRNMS 375
 QY 316 PRSPTSPGEGFVNDYGMFGAQTADADKTWTHGNHYAPNGSLGAMHYESKFRNMS 375
 DB PRSPTSPGEGFVNDYGMFGAQTADADKTWTHGNHYAPNGSLGAMHYESKFRNMS 375
 QY 301 GYSDPRGAYVITFIPKSWNTAPDKVKQMP 331
 DB GYSDPRGAYVITFIPKSWNTAPDKVKQMP 406
 RESULT 11
 AAB97831
 ID AAB97831 standard; Protein; 407 AA.
 XX
 AC AAB97831;
 XX
 DT 09-AUG-2001 (first entry)
 XX
 DE S. mobaraense IPO 13819 transglutaminase protein SPQ ID NO.4.
 XX
 KM Streptococcus lactis cinamomum IPO 12852; Streptomyces; actinomyces;
 KM Streptococcus lactis mobaraense IPO 13819; transglutaminase.
 XX
 OS Streptococcus lactis mobaraensis.
 XX
 PN WO200129187-A1.
 XX

PD 26-APR-2001.
 XX
 PF 13-OCT-2000; 2000MO-JP07135.
 XX
 PR 18-OCT-1999; 99JP-0295649.
 XX
 PA (AJIN) AJINOMOTO CO INC.
 XX
 PI Taguchi S, Momose H;
 XX
 DR WPI; 2001-300330/31.
 DR N-PSDB; AAH20188.
 XX
 PT Streptomyces sp. carrying an actinomycete-derived gene and promoter for
 PT producing high yields of transglutaminase
 PS Disclosure; Page 33-36; 41pp; Japanese.
 CC The present invention describes a Streptomyces sp. containing a gene
 CC construct comprising actinomycete-derived transglutaminase gene and
 CC promoter. Also described are methods for producing pro-transglutaminase
 CC and active transglutaminase. The gene construct can be used in the
 CC production of large amounts of transglutaminase. The present sequence
 CC represents Streptococcus lactis mobaraense IPO 13819 transglutaminase,
 CC which is given in the exemplification of the present invention.
 CC
 XX SQ Sequence 407 AA;
 Query Match 100.0%; Score 1811; DB 22; Length 407;
 Best Local Similarity 100.0%; Pred. No. 3.8e-153;
 Matches 331; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 DSDDRVTPPAEPLDRMPDPRPSYGRATVNNYIRKQOVYSHRDGRKQMTTEOREWL 60
 DB DSDDRVTPPAEPLDRMPDPRPSYGRATVNNYIRKQOVYSHRDGRKQMTTEOREWL 136
 QY 61 SYCGVGVTVWNSGOYPTNRLAFSPDEDFKNEKXGPRSGETRAEFGVAKESFDEE 120
 DB SYCGVGVTVWNSGOYPTNRLAFSPDEDFKNEKXGPRSGETRAEFGVAKESFDEE 196
 QY 121 KGFORAREVAVNNRBALENHDSAYLDNLKKELANGDALARNDARSFFYSALNTPSF 180
 DB KGFORAREVAVNNRBALENHDSAYLDNLKKELANGDALARNDARSFFYSALNTPSF 256
 QY 197 KGFORAREVAVNNRBALENHDSAYLDNLKKELANGDALARNDARSFFYSALNTPSF 256
 DB KGFORAREVAVNNRBALENHDSAYLDNLKKELANGDALARNDARSFFYSALNTPSF 316
 QY 181 KERNGNNDPSRMKAVIYSKHPWSGQDRSSADKRYGDPDAPRPAOTGLVMSRDNI 240
 DB KERNGNNDPSRMKAVIYSKHPWSGQDRSSADKRYGDPDAPRPAOTGLVMSRDNI 316
 QY 241 PRSPTSPGEGFVNDYGMFGAQTADADKTWTHGNHYAPNGSLGAMHYESKFRNMS 300
 DB PRSPTSPGEGFVNDYGMFGAQTADADKTWTHGNHYAPNGSLGAMHYESKFRNMS 376
 QY 317 PRSPTSPGEGFVNDYGMFGAQTADADKTWTHGNHYAPNGSLGAMHYESKFRNMS 376
 DB PRSPTSPGEGFVNDYGMFGAQTADADKTWTHGNHYAPNGSLGAMHYESKFRNMS 376
 QY 301 GYSDPRGAYVITFIPKSWNTAPDKVKQMP 331
 DB GYSDPRGAYVITFIPKSWNTAPDKVKQMP 407
 RESULT 12
 AAB81166
 ID AAB81166 standard; Protein; 407 AA.
 XX
 AC AAB81166;
 XX
 DT 13-JUL-2001 (first entry)
 XX
 DE Prepro-transglutaminase amino acid sequence.
 XX
 KM Corynebacterium bacteria; transglutaminase; food processing.
 XX
 OS Streptococcus lactis mobaraense.
 XX
 PN WO200123591-A1.
 XX

PD 05-APR-2001.
 XX
 XX 29-SEP-2000; 2000MO-JP06780.
 XX
 XX 30-SEP-1999; 99JP-0280098.
 PR 28-JUN-2000; 2000JP-0194043.
 XX
 PA (AJIN) AJINOMOTO CO INC.
 XX
 XX Kikuchi Y, Date M, Umezawa Y, Yokoyama K, Matsui H;
 XX
 DR WPI; 2001-266172/27.
 DR N-PSDB; AAF86283.
 XX
 XX Efficient secretory production of foreign proteins e.g.
 PT transglutaminase employing transformant coryneform bacterium, simply on
 PT industrial scale with direct recovery for use in food processing and
 PT pharmaceutical industry -
 XX
 XX Example 1; Page 90-93; 151pp; Japanese.
 XX
 CC This invention relates to a process for the production of a foreign
 CC secretory protein through the construction of a recombinant coryneform
 CC bacterium. The coryneform bacterium is transformed with an expression
 CC construct in which DNA encoding a target foreign protein pro-structure is
 CC ligated to the downstream region of DNA encoding the signal peptide
 CC domain of a coryneform bacterial protein. Following transformation with
 CC the vector, the bacterium is cultured, and the pro-peptide cleaved from
 CC the expressed protein. Transglutaminases produced using this process are
 CC useful in the food processing and pharmaceutical industries. The present
 CC sequence prepro-transglutaminase. The protein is used in an example
 CC illustrating the method of the invention.
 CC
 SQ Sequence 407 AA;
 Query Match 100.0%; Score 1811; DB 22; Length 407;
 Best Local Similarity 100.0%; Pred. No. 3.8e-153;
 Matches 331; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 DSDDRVTPPAEPLDRMPDPRPSYGRATVNNYIRKQOYSHRGRKQMTTEQRRL 60
 DB 77 DSDDRVTPPAEPLDRMPDPRPSYGRATVNNYIRKQOYSHRGRKQMTTEQRRL 136
 QY 61 SYGCVGTWVNSGOYPTNRLAFASPEDRFKNELKNGRPSGSTRAPFGYAKESPDEE 120
 DB 137 SYGCVGTWVNSGOYPTNRLAFASPEDRFKNELKNGRPSGSTRAPFGYAKESPDEE 196
 QY 122 KGFQAREVASVNNRALENNAHDESAVLNKKELANGNDALRNEDARSPFYSALRNTPSF 180
 DB 197 KGFQAREVASVNNRALENNAHDESAVLNKKELANGNDALRNEDARSPFYSALRNTPSF 256
 QY 181 KERNGNHDPBRMKAVIYSKHFWNSQDSSSSADKRYGDDPARPAPGGLVDSRRDNI 240
 DB 257 KERNGNHDPBRMKAVIYSKHFWNSQDSSSSADKRYGDDPARPAPGGLVDSRRDNI 316
 QY 241 PRSPFSPGEGFVNDYGMFGAQTEDADAKTWTGHNHYHAHPNSLGAMHVFYSKFRWSE 300
 DB 317 PRSPFSPGEGFVNDYGMFGAQTEDADAKTWTGHNHYHAHPNSLGAMHVFYSKFRWSE 376
 QY 301 GYSDFDRGAYVITFIKSNWNTAPDKYKQGW 331
 DB 377 GYSDFDRGAYVITFIKSNWNTAPDKYKQGW 407

XX
 XX Coryneform bacterium; signal peptide domain; food processing; medicine;
 KW cosmetic; transglutaminase; human epithelial growth factor.
 XX
 OS Streptococcus thermophilus mobaraense.
 XX
 XX WO200281694-A1.
 XX
 PD 17-OCT-2002.
 XX
 XX 27-MAR-2002; 2002WO-JP02978.
 PP 30-MAR-2001; 2001JP-0098808.
 XX
 XX (AJIN) AJINOMOTO CO INC.
 PA
 PI Kikuchi Y, Date M, Umezawa Y, Yokoyama K, Helma H, Matsui H;
 XX
 XX WPI; 2003-058550/05.
 DR N-PSDB; AEX10285.
 XX
 PT Secretion production of foreign proteins by culturing transformant
 PT coryneform bacteria, applicable in producing e.g. industrially-useful
 PT transglutaminase and human epithelial growth factor -
 XX
 XX Disclosure; Page 68-71; 117pp; Japanese.
 XX
 CC The invention relates to a method for producing a foreign protein by
 CC culturing a mutant strain of a coryneform bacterium that contains an
 CC expression gene construct to secrete the foreign protein, followed by
 CC recovery of the produced foreign protein. The gene construct is obtained
 CC by ligation of a nucleic acid encoding a signal peptide domain
 CC originating from a coryneform bacterium, to downstream of a functioning
 CC promoter sequence in the coryneform bacterium and also by ligation of a
 CC nucleic acid encoding a foreign protein, to downstream of a nucleic acid
 CC sequence encoding the signal peptide. The method is useful for the
 CC production of a foreign protein which is applicable in producing e.g.
 CC industrially-useful transglutaminase and human epithelial growth factor
 CC for use in medicine, cosmetics and food processing. This sequence
 CC represents a foreign protein of the invention.
 CC
 SQ Sequence 407 AA;
 Query Match 100.0%; Score 1811; DB 24; Length 407;
 Best Local Similarity 100.0%; Pred. No. 3.8e-153;
 Matches 331; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 DSDDRVTPPAEPLDRMPDPRPSYGRATVNNYIRKQOYSHRGRKQMTTEQRRL 60
 DB 77 DSDDRVTPPAEPLDRMPDPRPSYGRATVNNYIRKQOYSHRGRKQMTTEQRRL 136
 QY 61 SYGCVGTWVNSGOYPTNRLAFASPEDRFKNELKNGRPSGSTRAPFGYAKESPDEE 120
 DB 137 SYGCVGTWVNSGOYPTNRLAFASPEDRFKNELKNGRPSGSTRAPFGYAKESPDEE 196
 QY 122 KGFQAREVASVNNRALENNAHDESAVLNKKELANGNDALRNEDARSPFYSALRNTPSF 180
 DB 197 KGFQAREVASVNNRALENNAHDESAVLNKKELANGNDALRNEDARSPFYSALRNTPSF 256
 QY 181 KERNGNHDPBRMKAVIYSKHFWNSQDSSSSADKRYGDDPARPAPGGLVDSRRDNI 240
 DB 257 KERNGNHDPBRMKAVIYSKHFWNSQDSSSSADKRYGDDPARPAPGGLVDSRRDNI 316
 QY 241 PRSPFSPGEGFVNDYGMFGAQTEDADAKTWTGHNHYHAHPNSLGAMHVFYSKFRWSE 300
 DB 317 PRSPFSPGEGFVNDYGMFGAQTEDADAKTWTGHNHYHAHPNSLGAMHVFYSKFRWSE 376
 QY 301 GYSDFDRGAYVITFIKSNWNTAPDKYKQGW 331
 DB 377 GYSDFDRGAYVITFIKSNWNTAPDKYKQGW 407

RESULT 14

AA49049
ID AA49049 standard; Protein: 408 AA.
XX
AC AA49049;
XX
DT 10-MAR-2003 (updated)
DT 20-SEP-1994 (first entry)
XX
DE BTG-contg. sequence.
XX
KW Bacterial transglutaminase; BTG; expression;
KW active; inactive; inclusion body.
XX
OS Escherichia coli.
OS Synthetic.
XX
PN JP06030771-A.
XX
PD 08-FEB-1994.
XX
PF 14-JUL-1992; 92JP-0187038.
XX
PR 14-JUL-1992; 92JP-0187038.
XX
PA (AJIN) AJINOMOTO KK.
XX
DR WPI; 1994-079294/10.
DR N-PSDB; AAQ55984.
XX
PT Prep'n. of bacterial trans-glutaminase in large amts. - by
PT expression of fusion protein in E. coli bacterial
PT trans-glutaminase
XX
PS Disclosure; Page 8-10; 13pp; Japanese.
XX
CC A fused protein contains amino acids 16-346 of BTG (AAQ55983)
CC and a hydrophilic peptide at the amino terminal.
CC Expression of DNA encoding this protein in E. coli allow large
CC scale prodn. of BTG. An active BTG can be prep'd. from the
CC inactive fused protein inclusion body.
CC (updated on 10-MAR-2003 to add missing OS field.)
XX
XX
SQ Sequence 408 AA;
Query Match 100.0%; Score 1811; DB 15; Length 408;
Best Local Similarity 100.0%; Pred. No. 3,9e-153;
Matches 331; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 DSDDRVTPPAEPLDRPDPYRPSYGRATVNNYIRKQOYVSHDRKQOMTEBQREWL 60
DB 78 DSDDRVTPPAEPLDRPDPYRPSYGRATVNNYIRKQOYVSHDRKQOMTEBQREWL 137
QY 61 SYGCVGVTVNWSGOYPTNRLAFASFEDEDFKNEIKNGRPRSGETSAEFEGRVAKESFDEE 120
DB 138 SYGCVGVTVNWSGOYPTNRLAFASFEDEDFKNEIKNGRPRSGETSAEFEGRVAKESFDEE 197
QY 121 KGFQRAEVAIVMRALENAHDESAIYDNLKKELANGNDALRNEDARSPYSALRNTPEF 180
DB 198 KGFQRAEVAIVMRALENAHDESAIYDNLKKELANGNDALRNEDARSPYSALRNTPEF 257
QY 181 KERNGNHDSRMKAVIYSKHFWSGDRSSADKRYGDPDAFRAPGTGLVMSDRNT 240
DB 258 KERNGNHDSRMKAVIYSKHFWSGDRSSADKRYGDPDAFRAPGTGLVMSDRNT 317
QY 241 PRSPTSPGSGFVNFVDYMGFGAQTEDADKTWTHGNHYAPGSLGAMHYVESKFNWSE 300
DB 318 PRSPTSPGSGFVNFVDYMGFGAQTEDADKTWTHGNHYAPGSLGAMHYVESKFNWSE 377
QY 301 GYSPDNGAVITFIKSNRTLPDKVYQGNP 331
DB 378 GYSPDNGAVITFIKSNRTLPDKVYQGNP 408

RESULT 15
AA49049
ID AA49049 standard; Protein: 331 AA.
XX
AC AA49049;
XX
DT 25-MAR-2003 (updated)
DT 09-OCT-1992 (first entry)
XX
DE Transglutaminase (expressed in E. coli).
XX
KW BTG; acyl rearrangement; deamination; yeast; actinomycetes.
XX
OS Synthetic.
XX
PN EP481504-A.
XX
PD 22-APR-1992.
XX
PF 18-OCT-1991; 91EP-0117813.
XX
PR 19-OCT-1990; 90JP-0282566.
XX
PA (AMANO) AMANO PHARM KK.
XX
PA (AJIN) AJINOMOTO KK.
XX
PI Ando K, Arafuka S, Koikeda S, Matsui H, Takagi H, Washizu K;
XX
DR WPI; 1992-133808/17.
DR N-PSDB; AAQ24207.
XX
PT DNA fragment encoding trans-glutaminase - is inserted into
PT vector, e.g. pM1053-BTG, for protein expression
XX
XX
PS Disclosure; Page 3; 55pp; English.
XX
CC The protein sequence given has transglutaminase (BTG) activity. The
CC DNA encoding this protein has a base sequence which can be used
CC suitably in an expression system using E. coli or yeast as a host.
CC The base sequence can be compared to those given in AAQ24197 and
CC AAQ24200 which are derived from Actinomyces sp.
CC BTG catalyses an acyl rearrangement reaction of a gamma-carboxamide
CC gp. of glutamine. It introduces intra- or intermolecular formation
CC of epsilon-(gamma-Gln)-lys cross-linking when an epsilon-amino gp.
CC of a lys residue acts as an acyl receptor. When water acts as an
CC acyl acceptor the enzyme accelerates the conversion of Gln residues
CC to Glu residues by deamination.
CC The enzyme is used in the prodn. of gelled foods, gelled cosmetics,
CC yogurt, gelatin, cheese etc. It is also used in the prodn. of
CC thermally stable materials such as microcapsules and carriers of
CC immobilized enzymes. The DNA sequence given allows the prodn. of
CC BTG efficiently and in large quantity.
CC (Updated on 25-MAR-2003 to correct PA field.)
XX
XX
SQ Sequence 331 AA;
Query Match 99.5%; Score 1802; DB 13; Length 331;
Best Local Similarity 99.7%; Pred. No. 1.8e-152;
Matches 330; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 DSDDRVTPPAEPLDRPDPYRPSYGRATVNNYIRKQOYVSHDRKQOMTEBQREWL 60
DB 1 DSDDRVTPPAEPLDRPDPYRPSYGRATVNNYIRKQOYVSHDRKQOMTEBQREWL 60
QY 61 SYGCVGVTVNWSGOYPTNRLAFASFEDEDFKNEIKNGRPRSGETSAEFEGRVAKESFDEE 120
DB 61 SYGCVGVTVNWSGOYPTNRLAFASFEDEDFKNEIKNGRPRSGETSAEFEGRVAKESFDEE 120
QY 121 KGFQRAEVAIVMRALENAHDESAIYDNLKKELANGNDALRNEDARSPYSALRNTPEF 180
DB 121 KGFQRAEVAIVMRALENAHDESAIYDNLKKELANGNDALRNEDARSPYSALRNTPEF 180
QY 181 KERNGNHDSRMKAVIYSKHFWSGDRSSADKRYGDPDAFRAPGTGLVMSDRNT 240

Db	181	KERNQGNHDPSSRMKAVIYSKHFWNSGQDRSSSADKKRYGDDAFRPAFGTGLVDMGRVNI	240
Qy	241	PRSPSPGEGFVNFDYGMFGAQTADADADKTWTHGNHYHAPNGSLGAMHYESKFRMWSE	300
Db	241	PRSPSPGEGFVNFDYGMFGAQTADADADKTWTHGNHYHAPNGSLGAMHYESKFRMWSE	300
Qy	301	GYSDFDRGAYVITFIPKSWNTAPDKVKQGWMP	331
Db	301	GYSDFDRGAYVITFIPKSWNTAPDKVKQGWMP	331

Search completed: September 22, 2003, 12:43:55
Job time : 46 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 22, 2003, 12:43:10 / Search time 20 Seconds
(without alignments)
700,245 Million cell updates/sec

Title: US-09-884-948-1

Perfect score: 1811

Sequence: 1 DSDRYTPPAPBLDRMPDY.....ITFIPKSMNTAPDKVQGW 331

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 328717 seqs, 42310858 residues

Total number of hits satisfying chosen parameters: 328717

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents, AA.*
1: /cgn2_6/ptodata/1/1aa/5A-COMB.pep.*
2: /cgn2_6/ptodata/1/1aa/5S-COMB.pep.*
3: /cgn2_6/ptodata/1/1aa/6A-COMB.pep.*
4: /cgn2_6/ptodata/1/1aa/6S-COMB.pep.*
5: /cgn2_6/ptodata/1/1aa/6CTUS-COMB.pep.*
6: /cgn2_6/ptodata/1/1aa/backfill1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1811	100.0	331	1	US-08-136-993-1
2	1811	100.0	331	3	US-09-109-063-1
3	1811	100.0	331	3	US-08-793-426A-3
4	1811	100.0	331	3	US-08-294-565-3
5	1811	100.0	331	4	US-09-448-910-1
6	1811	100.0	406	1	US-08-136-993-13
7	1465	80.9	331	3	US-08-793-426A-2
8	1465	80.9	331	3	US-09-294-565-2
9	591	32.6	126	3	US-08-793-426A-7
10	591	32.6	126	3	US-09-294-565-7
11	208	11.5	40	3	US-08-793-426A-8
12	208	11.5	40	3	US-09-294-565-8
13	189	10.4	52	3	US-08-793-426A-6
14	189	10.4	52	3	US-09-294-565-6
15	102	5.6	735	4	US-09-321-589-1
16	99	5.5	315	3	US-08-558-135-5
17	93.5	5.2	1274	4	US-08-252-991A-20386
18	93	5.1	734	4	US-09-532-594B-4
19	91.5	5.1	333	4	US-09-252-991A-28443
20	91	5.0	544	1	US-08-264-002-7
21	91	5.0	544	4	US-09-457-040B-15
22	91	5.0	844	4	US-09-252-991A-25037
23	90.5	5.0	367	3	US-08-041-889-30
24	90.5	5.0	367	4	US-09-417-264-30
25	90.5	5.0	537	4	US-09-252-991A-27024
26	89.5	4.9	935	4	US-09-914-259-25
27	89	4.9	735	4	US-09-252-991A-18657

28	88	4.9	10182	4	US-09-134-001C-3159	Sequence 3159, Ap
29	87.5	4.8	1128	1	US-08-111-939-2	Sequence 2, Appl
30	87.5	4.8	1128	4	US-09-641-741-30	Sequence 30, Appl
31	87.5	4.8	1128	4	US-09-060-482-8	Sequence 8, Appl
32	86.5	4.8	913	1	US-08-487-890A-5	Sequence 5, Appl
33	86.5	4.8	913	2	US-08-478-435-5	Sequence 5, Appl
34	86.5	4.8	913	2	US-08-337-483-5	Sequence 5, Appl
35	86.5	4.8	913	2	US-08-478-373-5	Sequence 5, Appl
36	86.5	4.8	913	3	US-08-474-671-5	Sequence 5, Appl
37	86.5	4.8	913	3	US-08-483-577A-5	Sequence 5, Appl
38	86.5	4.8	913	3	US-08-897-438-5	Sequence 5, Appl
39	86.5	4.8	913	4	US-08-637-654-5	Sequence 5, Appl
40	86.5	4.8	913	4	US-08-649-518-5	Sequence 5, Appl
41	85	4.7	778	4	US-09-198-452A-508	Sequence 508, Ap
42	84.5	4.7	427	4	US-09-134-001C-5143	Sequence 5143, Ap
43	84.5	4.7	625	3	US-09-360-197-15	Sequence 15, Appl
44	84.5	4.7	674	3	US-08-893-852A-1	Sequence 1, Appl
45	84.5	4.7	702	4	US-09-107-532A-6866	Sequence 6866, Ap

ALIGNMENTS

RESULT: 1
US-08-136-993-1
Sequence 1, Application US/08136993
Patent No. 5420025
GENERAL INFORMATION:
APPLICANT: Takagi, Hiroshi
APPLICANT: Arafaka, Shino
APPLICANT: Matsui, Hiroshi
APPLICANT: Matsuz, Kinya
APPLICANT: Ando, Keiichi
APPLICANT: Koike, Satoshi
TITLE OF INVENTION: Recombinant transglutaminase
NUMBER OF SEQUENCES: 22
CORRESPONDENCE ADDRESSES:
ADDRESSEE: Sughrue, Mion, Zimm, Macpeak & Seas
STREET: 2100 Pennsylvania Avenue
CITY: N.W.
STATE: Washington, D.C.
COUNTRY: U.S.A.
ZIP: 20037-3202
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.24
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/136,993
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/07/777,447
FILING DATE:
APPLICATION NUMBER: JP 2-282566
FILING DATE: 19-OCT-1990
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-293-7060
TELEFAX: 202-293-7860
TELEX: 6491103
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 331 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-136-993-1
Query Match 100.0%; Score 1811; DB 1; Length 331;
Best Local Similarity 100.0%; Pred. No. 9.2e-171;
Matches 331; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DSDDRVTPPAEPLDMPPRPSYGAETVANNYIRKQOYSHRDGRKQOMTEBQREML 60
DB 1 DSDDRVTPPAEPLDMPPRPSYGAETVANNYIRKQOYSHRDGRKQOMTEBQREML 60
QY 61 SYGCVGVTVWNSGQYPTNRLAFASPEDERFKNELKNGRPRSGETRAFAFERVAKESFDEE 120
DB 61 SYGCVGVTVWNSGQYPTNRLAFASPEDERFKNELKNGRPRSGETRAFAFERVAKESFDEE 120
QY 121 KGFQARREVASVNNRALKENAHDESAVLDNLKKELANGNDALRNEDASPFYSALRNTSPF 180
DB 121 KGFQARREVASVNNRALKENAHDESAVLDNLKKELANGNDALRNEDASPFYSALRNTSPF 180
QY 181 KERNGNNDPSRMKAVIYSKHFWSGDRSSADKRYGDPAPRPAFGTGLVDMGRDNI 240
DB 181 KERNGNNDPSRMKAVIYSKHFWSGDRSSADKRYGDPAPRPAFGTGLVDMGRDNI 240
QY 241 PRSPTSPGEGFVNFYGFAGQTEADADKTWTHGNHYHAENGLGAMHYESKFRNMS 300
DB 241 PRSPTSPGEGFVNFYGFAGQTEADADKTWTHGNHYHAENGLGAMHYESKFRNMS 300
QY 301 GYSDPDRGAVYITFIPKSNWITAPDKVKQGW 331
DB 301 GYSDPDRGAVYITFIPKSNWITAPDKVKQGW 331

RESULT 2

US-09-109-063-1
; Sequence 1, Application US/09109063
; Patent No. 6013498
; GENERAL INFORMATION:
; APPLICANT: YOKOYAMA, KEIICHI
; APPLICANT: NAKAMURA, NAEMI
; APPLICANT: MIWA, TETSUYA
; APPLICANT: SEGURO, KATSUYA
; TITLE OF INVENTION: PROCESS FOR PRODUCING MICROBIAL TRANSGLUTAMINASE
; FILE REFERENCE: 0010-0937-0
; CURRENT APPLICATION NUMBER: US/09/109,063
; CURRENT FILING DATE: 1998-07-02
; EARLIER APPLICATION NUMBER: JP 180010/1997
; EARLIER FILING DATE: 1997-07-04
; NUMBER OF SEQ ID NOS: 62
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
; LENGTH: 331
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial
; OTHER INFORMATION: Sequence: TRANSGLUTAMINASE
US-09-109-063-1

Query Match 100.0%; Score 1811; DB 3; Length 331;
Best Local Similarity 100.0%; Pred. No. 9.2e-171;
Matches 331; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DSDDRVTPPAEPLDMPPRPSYGAETVANNYIRKQOYSHRDGRKQOMTEBQREML 60
DB 1 DSDDRVTPPAEPLDMPPRPSYGAETVANNYIRKQOYSHRDGRKQOMTEBQREML 60
QY 61 SYGCVGVTVWNSGQYPTNRLAFASPEDERFKNELKNGRPRSGETRAFAFERVAKESFDEE 120
DB 61 SYGCVGVTVWNSGQYPTNRLAFASPEDERFKNELKNGRPRSGETRAFAFERVAKESFDEE 120
QY 121 KGFQARREVASVNNRALKENAHDESAVLDNLKKELANGNDALRNEDASPFYSALRNTSPF 180
DB 121 KGFQARREVASVNNRALKENAHDESAVLDNLKKELANGNDALRNEDASPFYSALRNTSPF 180
QY 181 KERNGNNDPSRMKAVIYSKHFWSGDRSSADKRYGDPAPRPAFGTGLVDMGRDNI 240
DB 181 KERNGNNDPSRMKAVIYSKHFWSGDRSSADKRYGDPAPRPAFGTGLVDMGRDNI 240
QY 241 PRSPTSPGEGFVNFYGFAGQTEADADKTWTHGNHYHAENGLGAMHYESKFRNMS 300
DB 241 PRSPTSPGEGFVNFYGFAGQTEADADKTWTHGNHYHAENGLGAMHYESKFRNMS 300

DB 241 PRSPTSPGEGFVNFYGFAGQTEADADKTWTHGNHYHAENGLGAMHYESKFRNMS 300
QY 301 GYSDPDRGAVYITFIPKSNWITAPDKVKQGW 331
DB 301 GYSDPDRGAVYITFIPKSNWITAPDKVKQGW 331

RESULT 3

US-08-793-426A-3
; Sequence 3, Application US/08793426A
; Patent No. 6100053
; GENERAL INFORMATION:
; APPLICANT: Bech, Lisbeth
; APPLICANT: No. 6100053revang, Iben
; APPLICANT: Halkier, Torben
; APPLICANT: Rasmussen, Grethe
; APPLICANT: Schafer, Thomas
; TITLE OF INVENTION: Microbial Transglutaminases, Their
; TITLE OF INVENTION: Production And Use
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: No. 61000530 No. 6100053disk of No. 6100053th America, Inc.
; STREET: 405 Lexington Avenue
; CITY: New York
; STATE: NY
; COUNTRY: USA
; ZIP: 10174
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/793,426A
; FILING DATE: 25-FEB-1997
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Rozek, Carol E.
; REGISTRATION NUMBER: 36,993
; REFERENCE/DOCKET NUMBER: 4211.204-US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-867-0123
; TELEFAX: 212-878-9655
; TELEX:
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 331 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-793-426A-3

Query Match 100.0%; Score 1811; DB 3; Length 331;
Best Local Similarity 100.0%; Pred. No. 9.2e-171;
Matches 331; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DSDDRVTPPAEPLDMPPRPSYGAETVANNYIRKQOYSHRDGRKQOMTEBQREML 60
DB 1 DSDDRVTPPAEPLDMPPRPSYGAETVANNYIRKQOYSHRDGRKQOMTEBQREML 60
QY 61 SYGCVGVTVWNSGQYPTNRLAFASPEDERFKNELKNGRPRSGETRAFAFERVAKESFDEE 120
DB 61 SYGCVGVTVWNSGQYPTNRLAFASPEDERFKNELKNGRPRSGETRAFAFERVAKESFDEE 120
QY 121 KGFQARREVASVNNRALKENAHDESAVLDNLKKELANGNDALRNEDASPFYSALRNTSPF 180
DB 121 KGFQARREVASVNNRALKENAHDESAVLDNLKKELANGNDALRNEDASPFYSALRNTSPF 180
QY 181 KERNGNNDPSRMKAVIYSKHFWSGDRSSADKRYGDPAPRPAFGTGLVDMGRDNI 240
DB 181 KERNGNNDPSRMKAVIYSKHFWSGDRSSADKRYGDPAPRPAFGTGLVDMGRDNI 240
QY 241 PRSPTSPGEGFVNFYGFAGQTEADADKTWTHGNHYHAENGLGAMHYESKFRNMS 300
DB 241 PRSPTSPGEGFVNFYGFAGQTEADADKTWTHGNHYHAENGLGAMHYESKFRNMS 300

QY 241 PRSPSPGEGFVNFYDGMFGAQTEADADKTWTHGNHYHAPNGSLGAMHYESKFRNMSE 300
DB 241 PRSPSPGEGFVNFYDGMFGAQTEADADKTWTHGNHYHAPNGSLGAMHYESKFRNMSE 300
QY 301 GYSDFDRGAYVITFIKSWNTAPDKVKQGW 331
DB 301 GYSDFDRGAYVITFIKSWNTAPDKVKQGW 331

RESULT 4

US-09-294-565-3
; Sequence 3, Application US/09294565
; Patent No. 6190879

GENERAL INFORMATION:

APPLICANT: Sechl, Lisbeth
APPLICANT: No. 6190879;revang, Iben
APPLICANT: Halkier, Torden
APPLICANT: Rasmussen, Grethe
APPLICANT: Schafer, Thomas
APPLICANT: Andersen, Jens
TITLE OF INVENTION: Microbial Transglutaminases, Their
TITLE OF INVENTION: Production And Use
NUMBER OF SEQUENCES: 10
CORRESPONDENCE ADDRESS:
ADDRESSER: No. 6190879; No. 6190879;disk of No. 6190879;ch America, Inc.
STREET: 405 Lexington Avenue
CITY: New York
STATE: NY
COUNTRY: USA
ZIP: 10174
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/294,565
FILING DATE: 19-APR-1999
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Green, Reza
REGISTRATION NUMBER: 38,475
REFERENCE/DOCKET NUMBER: 4211,224-US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212-867-0123
TELEFAX: 212-878-9655
TELEX:
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 331 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-09-294-565-3

Query Match

Best Local Similarity 100.0%; Score 1811; DB 3; Length 331;
Matches 331; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DSDDRVTPPAEPLDRMPDPYRPSYGRATVYNNYIRKQOYVSHDGRKQOMTEQREML 60
DB 1 DSDDRVTPPAEPLDRMPDPYRPSYGRATVYNNYIRKQOYVSHDGRKQOMTEQREML 60
QY 61 SYGCVGVTVNSGOYPTNRLAFASPEDERFKNELKNGRPSGGETRAEFEGRAKESFDEE 120
DB 61 SYGCVGVTVNSGOYPTNRLAFASPEDERFKNELKNGRPSGGETRAEFEGRAKESFDEE 120
QY 121 KGFQARREVASVMMRALENADHDSAYLDNLKKELANGNDALRNEDARSPFYALNTPSF 180
DB 121 KGFQARREVASVMMRALENADHDSAYLDNLKKELANGNDALRNEDARSPFYALNTPSF 180

QY 181 KERNGNNDPSRMKAVIYS/HFWSGODRSSADKKYGDPAEPAPGGLVDMSPDRNI 240
DB 181 KERNGNNDPSRMKAVIYS/HFWSGODRSSADKKYGDPAEPAPGGLVDMSPDRNI 240
QY 241 PRSPSPGEGFVNFYDGMFGAQTEADADKTWTHGNHYHAPNGSLGAMHYESKFRNMSE 300
DB 241 PRSPSPGEGFVNFYDGMFGAQTEADADKTWTHGNHYHAPNGSLGAMHYESKFRNMSE 300
QY 301 GYSDFDRGAYVITFIKSWNTAPDKVKQGW 331
DB 301 GYSDFDRGAYVITFIKSWNTAPDKVKQGW 331

RESULT 5

US-09-448-310-1
; Sequence 1, Application US/09448310
; Patent No. 6538122

GENERAL INFORMATION:

APPLICANT: YOKOYAMA, KEIICHI
APPLICANT: NAKAMURA, NAMI
APPLICANT: MIWA, TETSUYA
APPLICANT: SEGRO, KATSUYA
TITLE OF INVENTION: PROCESS FOR PRODUCING MICROBIAL TRANSGLUTAMINASE
FILE REFERENCE: 0010-0937-0
CURRENT APPLICATION NUMBER: US/09/448,310
CURRENT FILING DATE: 1999-11-24
PRIOR APPLICATION NUMBER: 09/109,063
PRIOR FILING DATE: 1998-07-02
NUMBER OF SEQ ID NOS: 62
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 1
LENGTH: 331
TYPE: PRT
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Description of Artificial
OTHER INFORMATION: Sequence:TRANSGLUTAMINASE
US-09-448-310-1

Query Match

Best Local Similarity 100.0%; Score 1811; DB 4; Length 331;
Matches 331; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DSDDRVTPPAEPLDRMPDPYRPSYGRATVYNNYIRKQOYVSHDGRKQOMTEQREML 60
DB 1 DSDDRVTPPAEPLDRMPDPYRPSYGRATVYNNYIRKQOYVSHDGRKQOMTEQREML 60
QY 61 SYGCVGVTVNSGOYPTNRLAFASPEDERFKNELKNGRPSGGETRAEFEGRAKESFDEE 120
DB 61 SYGCVGVTVNSGOYPTNRLAFASPEDERFKNELKNGRPSGGETRAEFEGRAKESFDEE 120
QY 121 KGFQARREVASVMMRALENADHDSAYLDNLKKELANGNDALRNEDARSPFYALNTPSF 180
DB 121 KGFQARREVASVMMRALENADHDSAYLDNLKKELANGNDALRNEDARSPFYALNTPSF 180
QY 181 KERNGNNDPSRMKAVIYS/HFWSGODRSSADKKYGDPAEPAPGGLVDMSPDRNI 240
DB 181 KERNGNNDPSRMKAVIYS/HFWSGODRSSADKKYGDPAEPAPGGLVDMSPDRNI 240
QY 241 PRSPSPGEGFVNFYDGMFGAQTEADADKTWTHGNHYHAPNGSLGAMHYESKFRNMSE 300
DB 241 PRSPSPGEGFVNFYDGMFGAQTEADADKTWTHGNHYHAPNGSLGAMHYESKFRNMSE 300
QY 301 GYSDFDRGAYVITFIKSWNTAPDKVKQGW 331
DB 301 GYSDFDRGAYVITFIKSWNTAPDKVKQGW 331

RESULT 6

US-08-136-993-13
; Sequence 13, Application US/08136993
; Patent No. 5420025

GENERAL INFORMATION:

APPLICANT: Takagi, Hiroshi
APPLICANT: Araioka, Shino
APPLICANT: Matsui, Hiroshi
APPLICANT: Washizu, Kinya
APPLICANT: Ando, Keiichi
APPLICANT: Koikeda, Satoshi
TITLE OF INVENTION: Recombinant transglutaminase
NUMBER OF SEQUENCES: 22
CORRESPONDENCE ADDRESS:
ADDRESSEE: Sughrue, Mion, Zimm, Macpeak & Seas
STREET: 2100 Pennsylvania Avenue
CITY: N.W.
STATE: Washington, D.C.
COUNTRY: U.S.A.
ZIP: 20037-3202
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.24
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/136,993
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/07/777,447
FILING DATE:
APPLICATION NUMBER: JP 2-282566
FILING DATE: 19-OCT-1990
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-293-7060
TELEFAX: 202-293-7860
TELEX: 6491103
INFORMATION FOR SEQ ID NO: 13:
SEQUENCE CHARACTERISTICS:
LENGTH: 406 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-136-993-13

Query Match 100.0%; Score 1811; DB 1; Length 406;
Best Local Similarity 100.0%; Pred. No. 1.2e-170;
Matches 331; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 1 DSDRVTTPAPPLRMDPFRPSYGRATVYNNYIRKMQQVYSHRDGRKQWTEBQREWL 60
76 DSDRVTTPAPPLRMDPFRPSYGRATVYNNYIRKMQQVYSHRDGRKQWTEBQREWL 135
Qy 61 SYGCVGVTVWNSGQYFPNRLAFASPDDEFRKNEIKNGRPPSGETRAEFEGRYAKSPDEE 120
Db 136 SYGCVGVTVWNSGQYFPNRLAFASPDDEFRKNEIKNGRPPSGETRAEFEGRYAKSPDEE 195
Qy 121 KGFQRAREVASVNNRALENHADSAYLDNLKKEILANGNDALREEDARSPFYALRNTPSF 180
Db 196 KGFQRAREVASVNNRALENHADSAYLDNLKKEILANGNDALREEDARSPFYALRNTPSF 255
Qy 181 KERNGGNHDPBRKMAVYISKHFWSGQDSSSADRKKYGDPAFRPAPGTGLVDMKDRNIP 240
Db 256 KERNGGNHDPBRKMAVYISKHFWSGQDSSSADRKKYGDPAFRPAPGTGLVDMKDRNIP 315
Qy 241 PRSPTPSGEGVNFVDYGMFGAQTEDADKXTWTGNHNYHAPNGSLGAMHYESKFRNWS 300
Db 316 PRSPTPSGEGVNFVDYGMFGAQTEDADKXTWTGNHNYHAPNGSLGAMHYESKFRNWS 375
Qy 301 GYSDFDRGAYVITFIPKSNMTAPDKVKQGW 331
Db 376 GYSDFDRGAYVITFIPKSNMTAPDKVKQGW 406

RESULT 7
US-08-793-426A-2
; Sequence 2, Application US/08793426A

Patent No. 6100053
GENERAL INFORMATION:
APPLICANT: Bech, Lisbeth
APPLICANT: No. 6100053revang, Iben
APPLICANT: Halkier, Torben
APPLICANT: Rasmussen, Grethe
APPLICANT: Schaffer, Thomas
APPLICANT: Andersen, Jens
TITLE OF INVENTION: Microbial Transglutaminases, Their
NUMBER OF SEQUENCES: 10
CORRESPONDENCE ADDRESS:
ADDRESSEE: No. 61000530 No. 6100053disk of No. 6100053th America, Inc.
STREET: 405 Lexington Avenue
CITY: New York
STATE: NY
COUNTRY: USA
ZIP: 10174
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/793,426A
FILING DATE: 25-FEB-1997
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Rozek, Carol E.
REGISTRATION NUMBER: 36,993
REFERENCE/DOCKET NUMBER: 4211.204-US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212-867-0123
TELEFAX: 212-878-9655
TELEX:
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 331 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-793-426A-2

Query Match 80.9%; Score 1465; DB 3; Length 331;
Best Local Similarity 79.3%; Pred. No. 1.4e-136;
Matches 261; Conservative 34; Mismatches 34; Indels 0; Gaps 0;

Qy 2 SDRVTTPAPPLRMDPFRPSYGRATVYNNYIRKMQQVYSHRDGRKQWTEBQREWL 61
Db 2 ADRVTTPAPPLRMDPFRPSYGRATVYNNYIRKMQQVYSHRDGRKQWTEBQREWL 61
Qy 62 YGCVGVTVWNSGQYFPNRLAFASPDDEFRKNEIKNGRPPSGETRAEFEGRYAKSPDEE 121
Db 62 YGCVGVTVWNSGQYFPNRLAFASPDDEFRKNEIKNGRPPSGETRAEFEGRYAKSPDEE 121
Qy 122 GFQRAREVASVNNRALENHADSAYLDNLKKEILANGNDALREEDARSPFYALRNTPSF 181
Db 122 GFQRAREVASVNNRALENHADSAYLDNLKKEILANGNDALREEDARSPFYALRNTPSF 181
Qy 182 ERNGGNHDPBRKMAVYISKHFWSGQDSSSADRKKYGDPAFRPAPGTGLVDMKDRNIP 241
Db 182 ERNGGNHDPBRKMAVYISKHFWSGQDSSSADRKKYGDPAFRPAPGTGLVDMKDRNIP 241
Qy 242 RSPSPSGEGVNFVDYGMFGAQTEDADKXTWTGNHNYHAPNGSLGAMHYESKFRNWS 301
Db 242 RSPSPSGEGVNFVDYGMFGAQTEDADKXTWTGNHNYHAPNGSLGAMHYESKFRNWS 301
Qy 302 YSDFDRGAYVITFIPKSNMTAPDKVKQGW 330
Db 302 YSDFDRGAYVITFIPKSNMTAPDKVKQGW 330

RESULT 8

US-09-294-565-2

Sequence 2, Application US/09294565
 Patent No. 6190879
 GENERAL INFORMATION:
 APPLICANT: Bech, Lisbeth
 APPLICANT: No. 6190879revang, Iben
 APPLICANT: Halkier, Torben
 APPLICANT: Rasmussen, Grethe
 APPLICANT: Schafer, Thomas
 APPLICANT: Andersen, Jens
 TITLE OF INVENTION: Microbial Transglutaminases, Their
 NUMBER OF INVENTIONS: Production And Use
 NUMBER OF SEQUENCES: 10
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: No. 61908790 No. 6190879disk of No. 6190879ch America, Inc.
 STREET: 405 Lexington Avenue
 CITY: New York
 STATE: NY
 COUNTRY: USA
 ZIP: 10174
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Diskette
 OPERATING SYSTEM: DOS
 SOFTWARE: FastSeq for Windows Version 2.0
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/09/294,565
 FILING DATE: 19-APR-1999
 CLASSIFICATION:
 ATTORNEY/AGENT INFORMATION:
 NAME: Green, Reza
 REGISTRATION NUMBER: 38,475
 REFERENCE/DOCKET NUMBER: 4211.224-US
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 212-867-0123
 TELEFAX: 212-878-9655
 TELEX:
 INFORMATION FOR SEQ ID NO: 2:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 331 amino acids
 TYPE: amino acid
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 US-09-294-565-2

Query Match 80.9%; Score 1465; DB 3; Length 331;
 Best Local Similarity 79.3%; Pred. No. 1.4e-136;
 Matches 261; Conservative 34; Mismatches 34; Indels 0; Gaps 0;

QY 2 SDRVTVPPEPLDNRNDEPRPSYGRATVYNNYIRKQOQVYSHRDRKQOMTEBQREWL 61
 DB 2 ADRVTVPPEPLDNRNDEPRPSYGRATVYNNYIRKQOQVYSHRDRKQOMTEBQREWL 61
 QY 62 YGVGVTVNWSGQVYTNRIAFASFDPRFKNLKXGRPSGERTRAEFERVAKSFDEEK 121
 DB 62 YGVGVTVNWSGQVYTNRIAFASFDPRFKNLKXGRPSGERTRAEFERVAKSFDEEK 121
 QY 122 GFORAEVAVSVNRRLLENADESAAYLDNKKELANGNDALNEDASPPYSALRNTPEFK 181
 DB 122 GFORAEVAVSVNRRLLENADESAAYLDNKKELANGNDALNEDASPPYSALRNTPEFK 181
 QY 182 ERKGAVHDSRKATVYSKGFHWSGDRSSSAKRYGPDPAFRAPGTGLVMSRDRIIP 241
 DB 182 ERKGAVHDSRKATVYSKGFHWSGDRSSSAKRYGPDPAFRAPGTGLVMSRDRIIP 241
 QY 242 RSPSTSGEGFVNFYDGMFGAQTEADADKTWTHGNHYHAPNGSLGAMHYESKFPNMSG 301
 DB 242 RSPSTSGEGFVNFYDGMFGAQTEADADKTWTHGNHYHAPNGSLGAMHYESKFPNMSG 301
 QY 302 YSDFDGAIVYTFIKSWNTAPDKVYKQW 330
 DB 302 YSDFDGAIVYTFIKSWNTAPDKVYKQW 330

RESULT 9

US-08-793-426A-7
 Sequence 7, Application US/08793426A
 Patent No. 6100053
 GENERAL INFORMATION:
 APPLICANT: Bech, Lisbeth
 APPLICANT: No. 6100053revang, Iben
 APPLICANT: Halkier, Torben
 APPLICANT: Rasmussen, Grethe
 APPLICANT: Schafer, Thomas
 APPLICANT: Andersen, Jens
 TITLE OF INVENTION: Microbial Transglutaminases, Their
 NUMBER OF INVENTIONS: Production And Use
 NUMBER OF SEQUENCES: 10
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: No. 61000530 No. 6100053disk of No. 6100053ch America, Inc.
 STREET: 405 Lexington Avenue
 CITY: New York
 STATE: NY
 COUNTRY: USA
 ZIP: 10174
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Diskette
 OPERATING SYSTEM: DOS
 SOFTWARE: FastSeq for Windows Version 2.0
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/793,426A
 FILING DATE: 25-FEB-1997
 CLASSIFICATION: 435
 ATTORNEY/AGENT INFORMATION:
 NAME: Rozek, Carol B.
 REGISTRATION NUMBER: 36,993
 REFERENCE/DOCKET NUMBER: 4211.204-US
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 212-867-0123
 TELEFAX: 212-878-9655
 TELEX:
 INFORMATION FOR SEQ ID NO: 7:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 126 amino acids
 TYPE: amino acid
 STRANDEDNESS:
 TOPOLOGY: linear
 MOLECULE TYPE: peptide
 HYPOTHEICAL: NO
 ANTI-SENSE: NO
 FRAGMENT TYPE: internal
 US-08-793-426A-7

Query Match 32.6%; Score 591; DB 3; Length 126;
 Best Local Similarity 82.5%; Pred. No. 8e-51;
 Matches 104; Conservative 8; Mismatches 14; Indels 0; Gaps 0;

QY 165 DARPPFSAIRNTPSRKERNGANDPSRKAVYYSKGFHWSGDRSSADRRKGDPAAR 224
 DB 1 DSRSSFSALRNTPSRKERNGANDPSRKAVYYSKGFHWSGDRSSADRRKGDPAAR 60
 QY 225 PAPGTGLVDSRDRIIPSPSPSGEFGVNFYDGMFGAQTEADADKTWTHGNHYHAPNGS 284
 DB 61 PDGTGLVDSRDRIIPSPSPSGEFGVNFYDGMFGAQTEADADKTWTHGNHYHAPNGS 120
 QY 285 LGANHV 290
 DB 121 MGPMBV 126

RESULT 10

US-09-294-565-7
 Sequence 7, Application US/09294565
 Patent No. 6190879
 GENERAL INFORMATION:

APPLICANT: Bech, Lisbeth
APPLICANT: No. 6190879revang, Iben
APPLICANT: Halkier, Torben
APPLICANT: Rasmussen, Grethe
APPLICANT: Schaffer, Thomas
APPLICANT: Andersen, Jens
TITLE OF INVENTION: Microbial Transglutaminases, Their
TITLE OF INVENTION: Production And Use
NUMBER OF SEQUENCES: 10
CORRESPONDENCE ADDRESS:
ADDRESS: No. 61908790 No. 6190879disk of No. 6190879th America, Inc.
STREET: 405 Lexington Avenue
CITY: New York
STATE: NY
COUNTRY: USA
ZIP: 10174
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/294,565
FILING DATE: 19-APR-1997
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Green, Reza
REGISTRATION NUMBER: 38,475
REFERENCE/DOCKET NUMBER: 4211.224-US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212-867-0123
TELEFAX: 212-878-9655
TELEX:
INFORMATION FOR SEQ ID NO: 7:
SEQUENCE CHARACTERISTICS:
LENGTH: 126 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
MOLECULE TYPE: peptide
HYPOTHETICAL: NO
ANTI-SENSE: NO
FRAGMENT TYPE: internal
US-09-294-565-7
Query Match 32.6%; Score 591; DB 3; Length 126;
Best Local Similarity 82.5%; Pred. No. 8e-51;
Matches 104; Conservative 8; Mismatches 14; Indels 0; Gaps 0;
QY 165 DARPPTYSALRNTPSKREKNGNDPSRKAVYISKHFWSGQDRSSSDKRYKDDPDAPR 224
DB 1 DRSSEFYSALRNTPSKREKNGNDPSRKAVYISKHFWSGQDRSSSDKRYKDDPDAPR 60
QY 225 PAPGTGLVDMRDNRNTPRSPTSGRGFPVNFYGFQAQTEADADADKTWTGHNHYHAAHNS 284
DB 61 PDGTGLVDMRDNRNTPRSPTSGRGFPVNFYGFQAQTEADADADKTWTGHNHYHABGG 120
QY 285 LGAMHV 290
DB 121 MGPMHV 126
RESULT 11
US-08-793-426A-8
Sequence 8, Application US/08793426A
Patent No. 6100053
GENERAL INFORMATION:
APPLICANT: Bech, Lisbeth
APPLICANT: No. 6100053revang, Iben
APPLICANT: Halkier, Torben
APPLICANT: Rasmussen, Grethe
APPLICANT: Schaffer, Thomas
APPLICANT: Andersen, Jens

TITLE OF INVENTION: Microbial Transglutaminases, Their
TITLE OF INVENTION: Production And Use
NUMBER OF SEQUENCES: 10
CORRESPONDENCE ADDRESS:
ADDRESS: No. 61000530 No. 6100053disk of No. 6100053th America, Inc.
STREET: 405 Lexington Avenue
CITY: New York
STATE: NY
COUNTRY: USA
ZIP: 10174
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/793,426A
FILING DATE: 25-FEB-1997
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Rozek, Carol E.
REGISTRATION NUMBER: 36,993
REFERENCE/DOCKET NUMBER: 4211.204-US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212-867-0123
TELEFAX: 212-878-9655
TELEX:
INFORMATION FOR SEQ ID NO: 8:
SEQUENCE CHARACTERISTICS:
LENGTH: 40 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
MOLECULE TYPE: peptide
HYPOTHETICAL: NO
ANTI-SENSE: NO
FRAGMENT TYPE: C-terminal
US-08-793-426A-8
Query Match 11.5%; Score 208; DB 3; Length 40;
Best Local Similarity 90.0%; Pred. No. 1e-13;
Matches 36; Conservative 2; Mismatches 2; Indels 0; Gaps 0;
QY 292 ESKFRWMSGYSDFDGAAYVITFIPKSWNTAPDKYKQGP 331
DB 1 ESKFRWMSAGYADFDRGAYVITFIPKSWNTAPDKYKQGP 40
RESULT 12
US-09-294-565-8
Sequence 8, Application US/09294565
Patent No. 6190879
GENERAL INFORMATION:
APPLICANT: Bech, Lisbeth
APPLICANT: No. 6190879revang, Iben
APPLICANT: Halkier, Torben
APPLICANT: Rasmussen, Grethe
APPLICANT: Schaffer, Thomas
APPLICANT: Andersen, Jens
TITLE OF INVENTION: Microbial Transglutaminases, Their
TITLE OF INVENTION: Production And Use
NUMBER OF SEQUENCES: 10
CORRESPONDENCE ADDRESS:
ADDRESS: No. 61908790 No. 6190879disk of No. 6190879th America, Inc.
STREET: 405 Lexington Avenue
CITY: New York
STATE: NY
COUNTRY: USA
ZIP: 10174
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS


```
SOFTWARE: FastSeq for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/294,565
FILING DATE: 19-APR-1999
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Green, Reza
REGISTRATION NUMBER: 38,475
REFERENCE/DOCKET NUMBER: 4211.224-US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212-867-0123
TELEFAX: 212-878-9655
TELEX:
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 40 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
MOLECULE TYPE: peptide
HYPOTHEICAL: NO
ANTI-SENSE: NO
FRAGMENT TYPE: C-terminal
US-09-294-565-8

Query Match          11.5%; Score 208; DB 3; Length 40;
Best Local Similarity 90.0%; Pred. No. 1e-13;
Matches 36; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY      292 ESKFRNWSRGISDPDRAGAYITIPKSNMTAPDKYKQGP 331
Db      1 ESKFRNWSAGYADPDGAYITIPKSNMTAPAEVKQGP 40

RESULT 13
US-08-793-426A-6
Sequence 6, Application US/08793426A
Patent No. 6100053
GENERAL INFORMATION:
APPLICANT: Bech, Lisbeth
APPLICANT: No. 6100053revang, Iben
APPLICANT: Halkier, Torben
APPLICANT: Rasmussen, Grethe
APPLICANT: Schaefer, Thomas
APPLICANT: Andersen, Jens
TITLE OF INVENTION: Microbial Transglutaminases, Their
NUMBER OF SEQUENCES: 10
CORRESPONDENCE ADDRESS:
ADDRESSER: No. 61000530 No. 6100053disk of No. 6100053ch America, Inc.
STREET: 405 Lexington Avenue
CITY: New York
STATE: NY
COUNTRY: USA
ZIP: 10174
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/793,426A
FILING DATE: 25-FEB-1997
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Rozek, Carol E.
REGISTRATION NUMBER: 36,993
REFERENCE/DOCKET NUMBER: 4211.204-US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212-867-0123
TELEFAX: 212-878-9655
TELEX:
INFORMATION FOR SEQ ID NO: 6:
```

```
SEQUENCE CHARACTERISTICS:
LENGTH: 52 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
MOLECULE TYPE: peptide
HYPOTHEICAL: NO
FRAGMENT TYPE: internal
US-08-793-426A-6

Query Match          10.4%; Score 189; DB 3; Length 52;
Best Local Similarity 70.6%; Pred. No. 1.1e-11;
Matches 36; Conservative 8; Mismatches 7; Indels 0; Gaps 0;

QY      106 AEFGRVAKESFDEKGFQARREYASVNMRALENHDSAYLDNKKELAN 156
Db      1 AEFGRVAKGFADRGKREYASVNMRALENHDSAYLDNKKELAN 51

RESULT 14
US-09-294-565-6
Sequence 6, Application US/09294565
Patent No. 6190879
GENERAL INFORMATION:
APPLICANT: Bech, Lisbeth
APPLICANT: No. 6190879revang, Iben
APPLICANT: Halkier, Torben
APPLICANT: Rasmussen, Grethe
APPLICANT: Schaefer, Thomas
APPLICANT: Andersen, Jens
TITLE OF INVENTION: Microbial Transglutaminases, Their
NUMBER OF SEQUENCES: 10
CORRESPONDENCE ADDRESS:
ADDRESSER: No. 61908790 No. 6190879disk of No. 6190879ch America, Inc.
STREET: 405 Lexington Avenue
CITY: New York
STATE: NY
COUNTRY: USA
ZIP: 10174
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/294,565
FILING DATE: 19-APR-1999
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Green, Reza
REGISTRATION NUMBER: 38,475
REFERENCE/DOCKET NUMBER: 4211.224-US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212-867-0123
TELEFAX: 212-878-9655
TELEX:
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 52 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
MOLECULE TYPE: peptide
HYPOTHEICAL: NO
FRAGMENT TYPE: internal
US-09-294-565-6

Query Match          10.4%; Score 189; DB 3; Length 52;
Best Local Similarity 70.6%; Pred. No. 1.1e-11;
Matches 36; Conservative 8; Mismatches 7; Indels 0; Gaps 0;

QY      106 AEFGRVAKESFDEKGFQARREYASVNMRALENHDSAYLDNKKELAN 156
```


GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 22, 2003, 12:45:10 ; Search time 62 Seconds

(without alignments)
794.898 Million cell updates/sec

Title: US-09-884-948-1

Perfect score: 1811

Sequence: 1 DSDRVTPEPAPLDRKMPDY.....ITFPMKMTAPDKVKGMP 331

Scoring table:

BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 556269 seqs, 148893369 residues

Total number of hits satisfying chosen parameters: 556269

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications AA.*
1: /cgn2_6/prodata/2/pubpaa/US07_PUBCOMB.pep.*
2: /cgn2_6/prodata/2/pubpaa/PCT_NEW_PUB.pep.*
3: /cgn2_6/prodata/2/pubpaa/US06_NEW_PUB.pep.*
4: /cgn2_6/prodata/2/pubpaa/US06_PUBCOMB.pep.*
5: /cgn2_6/prodata/2/pubpaa/US07_NEW_PUB.pep.*
6: /cgn2_6/prodata/2/pubpaa/PCTUS_PUBCOMB.pep.*
7: /cgn2_6/prodata/2/pubpaa/US08_NEW_PUB.pep.*
8: /cgn2_6/prodata/2/pubpaa/US08_PUBCOMB.pep.*
9: /cgn2_6/prodata/2/pubpaa/US09_PUBCOMB.pep.*
10: /cgn2_6/prodata/2/pubpaa/US09_PUBCOMB.pep.*
11: /cgn2_6/prodata/2/pubpaa/US09C_PUBCOMB.pep.*
12: /cgn2_6/prodata/2/pubpaa/US09_NEW_PUB.pep.*
13: /cgn2_6/prodata/2/pubpaa/US10_PUBCOMB.pep.*
14: /cgn2_6/prodata/2/pubpaa/US10_PUBCOMB.pep.*
15: /cgn2_6/prodata/2/pubpaa/US10C_PUBCOMB.pep.*
16: /cgn2_6/prodata/2/pubpaa/US10_NEW_PUB.pep.*
17: /cgn2_6/prodata/2/pubpaa/US60_NEW_PUB.pep.*
18: /cgn2_6/prodata/2/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1811	100.0	331	10 US-09-996-561-1	Sequence 1, Appl1
2	1811	100.0	331	10 US-09-884-948-1	Sequence 1, Appl1
3	1811	100.0	331	15 US-10-112-488-5	Sequence 5, Appl1
4	1811	100.0	332	9 US-09-892-864-2	Sequence 2, Appl1
5	1811	100.0	407	14 US-10-124-493-4	Sequence 4, Appl1
6	1811	100.0	407	15 US-10-112-488-13	Sequence 13, Appl1
7	1811	95.3	410	15 US-10-022-809-2	Sequence 2, Appl1
8	1487	82.1	330	15 US-10-112-488-43	Sequence 43, Appl1
9	1487	82.1	416	14 US-10-124-493-2	Sequence 2, Appl1
10	122.5	6.8	738	12 US-10-291-583-92	Sequence 92, Appl1
11	122.5	6.8	738	12 US-10-291-583-93	Sequence 93, Appl1
12	122.5	6.8	738	12 US-10-291-583-94	Sequence 94, Appl1
13	122	6.7	737	12 US-10-291-583-95	Sequence 2, Appl1
14	117	6.5	736	12 US-10-291-583-99	Sequence 99, Appl1
15	117	6.5	737	12 US-10-291-583-72	Sequence 72, Appl1

16	114.5	6.3	1217	12 US-10-311-406-2	Sequence 2, Appl1
17	114	6.3	735	12 US-10-291-583-68	Sequence 68, Appl1
18	114	6.3	735	12 US-10-291-583-69	Sequence 69, Appl1
19	113.5	6.3	729	12 US-10-291-583-111	Sequence 111, Appl1
20	113.5	6.3	733	12 US-10-291-583-60	Sequence 60, Appl1
21	113.5	6.3	733	12 US-10-291-583-62	Sequence 62, Appl1
22	113.5	6.3	738	12 US-10-291-583-95	Sequence 95, Appl1
23	112.5	6.2	733	12 US-10-291-583-61	Sequence 61, Appl1
24	112	6.2	733	12 US-10-291-583-86	Sequence 86, Appl1
25	112	6.2	736	12 US-10-291-583-96	Sequence 96, Appl1
26	112	6.2	736	12 US-10-291-583-97	Sequence 97, Appl1
27	112	6.2	736	12 US-10-291-583-98	Sequence 98, Appl1
28	111.5	6.2	729	12 US-10-291-583-109	Sequence 109, Appl1
29	111.5	6.2	729	12 US-10-291-583-110	Sequence 110, Appl1
30	111.5	6.2	738	12 US-10-291-583-78	Sequence 78, Appl1
31	111.5	6.2	738	12 US-10-291-583-80	Sequence 80, Appl1
32	111.5	6.2	738	12 US-10-291-583-81	Sequence 81, Appl1
33	111.5	6.2	738	12 US-10-291-583-82	Sequence 82, Appl1
34	111	6.1	728	12 US-10-291-583-103	Sequence 103, Appl1
35	111	6.1	728	12 US-10-291-583-104	Sequence 104, Appl1
36	110	6.1	728	12 US-10-291-583-108	Sequence 108, Appl1
37	110	6.1	736	12 US-10-291-583-71	Sequence 71, Appl1
38	109.5	6.0	738	12 US-10-291-583-83	Sequence 83, Appl1
39	109.5	6.0	738	12 US-10-291-583-84	Sequence 84, Appl1
40	109.5	6.0	738	12 US-10-291-583-85	Sequence 85, Appl1
41	109	6.0	728	12 US-10-291-583-101	Sequence 101, Appl1
42	108	6.0	728	12 US-10-291-583-102	Sequence 102, Appl1
43	108	6.0	728	12 US-10-291-583-105	Sequence 105, Appl1
44	108	6.0	733	12 US-10-291-583-90	Sequence 90, Appl1
45	108	6.0	1005	9 US-09-925-301-1335	Sequence 1335, Appl1

ALIGNMENTS

RESULT 1
US-09-996-561-1
; Sequence 1, Application US/09996561
; Patent No. US20020151703A1
; GENERAL INFORMATION:
; APPLICANT: YOKOYAMA, KEIICHI
; APPLICANT: YOKOYAMA, NAAMI
; APPLICANT: MIWA, TETSUYA
; APPLICANT: SESUGO, KATSUYA
; TITLE OF INVENTION: PROCESS FOR PRODUCING MICROBIAL TRANSGLUTAMINASE
; FILE REFERENCE: 0010-0937-0
; CURRENT APPLICATION NUMBER: US/09/996,561
; CURRENT FILING DATE: 2001-11-30
; PRIOR APPLICATION NUMBER: CURRENT APPLICATION NUMBER: US/09/448,310
; PRIOR FILING DATE: CURRENT FILING DATE: 1999-11-24
; PRIOR APPLICATION NUMBER: 09/109,063
; PRIOR FILING DATE: 1998-07-02
; NUMBER OF SEQ ID NOS: 62
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
; LENGTH: 331
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial
; OTHER INFORMATION: Sequence:TRANSGLUTAMINASE
US-09-996-561-1

Query Match 100.0%; Score 1811; DB 10; Length 331;

Best local Similarity 100.0%; Pred. No. 5; 5e-167; Mismatches 0; Indels 0; Gaps 0;

Matches 331; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DSDRVTPEPAPLDRKMPDYPSYGRLETIVNNYTRKQOYVSHRDRKQOMTEORENL 60
DB 1 DSDRVTPEPAPLDRKMPDYPSYGRLETIVNNYTRKQOYVSHRDRKQOMTEORENL 60
QY 61 SYGCVGTWNSGQVTRRLAFASFDREKLEKNGRREGTRAFBRVAKESFDEE 120

Db 61 SYGCVGVTWWSGQYPTNRLAFASFEDEDRFKNELKNGRPRSGETPAEFGGVAKESFDEE 120
QY 121 KGFQARAEVASVNNRALENAHDESAYLNDLKKELANGDNLREDAARSPFSALRNTPSF 180
Db 121 KGFQARAEVASVNNRALENAHDESAYLNDLKKELANGDNLREDAARSPFSALRNTPSF 180
QY 181 KERNGNHDPSRMKAVIYSKHFWSGQDRSSADRRKYGDPDAFRPAPGTGLVMSRDRI 240
Db 181 KERNGNHDPSRMKAVIYSKHFWSGQDRSSADRRKYGDPDAFRPAPGTGLVMSRDRI 240
QY 241 PRSPFSGEGFVNFDFYGMFGAQTEDADKTVWTHGNHYHAPNGSLGAMHYESKFRWSE 300
Db 241 PRSPFSGEGFVNFDFYGMFGAQTEDADKTVWTHGNHYHAPNGSLGAMHYESKFRWSE 300
QY 301 GYSDFDRGAVYITFIPIKSWNTAPDKVXKQGP 331
Db 301 GYSDFDRGAVYITFIPIKSWNTAPDKVXKQGP 331

RESULT 2
US-09-884-948-1
; Sequence 1, Application US/09884948
; Patent No. US20020173021A1
; GENERAL INFORMATION:
; APPLICANT: YOKOYAMA, KEIICHI
; APPLICANT: NAKAMURA, NAHI
; APPLICANT: MIWA, TETSUYA
; APPLICANT: SEGURO, KATSUYA
; TITLE OF INVENTION: PROCESS FOR PRODUCING MICROBIAL TRANSGLUTAMINASE
; FILE REFERENCE: 0010-0937-0
; CURRENT APPLICATION NUMBER: US/09/884,948
; CURRENT FILING DATE: 2001-06-21
; PRIOR FILING DATE: 1999-11-24
; PRIOR APPLICATION NUMBER: 09/448,310
; NUMBER OF SEQ ID NOS: 62
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
; LENGTH: 331
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial
; OTHER INFORMATION: Sequence: TRANSGLUTAMINASE
US-09-884-948-1

Query Match 100.0%; Score 1811; DB 10; Length 331;
Best Local Similarity 100.0%; Pred. No. 5, 5e-167;
Matches 331; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DSDRAVTPPAEPLDRLPDPYRPSYGRATVYNNYIRKQOYVSHRDGRKQOMTEBOEWL 60
Db 1 DSDRAVTPPAEPLDRLPDPYRPSYGRATVYNNYIRKQOYVSHRDGRKQOMTEBOEWL 60
QY 61 SYGCVGVTWWSGQYPTNRLAFASFEDEDRFKNELKNGRPRSGETPAEFGGVAKESFDEE 120
Db 61 SYGCVGVTWWSGQYPTNRLAFASFEDEDRFKNELKNGRPRSGETPAEFGGVAKESFDEE 120
QY 121 KGFQARAEVASVNNRALENAHDESAYLNDLKKELANGDNLREDAARSPFSALRNTPSF 180
Db 121 KGFQARAEVASVNNRALENAHDESAYLNDLKKELANGDNLREDAARSPFSALRNTPSF 180
QY 181 KERNGNHDPSRMKAVIYSKHFWSGQDRSSADRRKYGDPDAFRPAPGTGLVMSRDRI 240
Db 181 KERNGNHDPSRMKAVIYSKHFWSGQDRSSADRRKYGDPDAFRPAPGTGLVMSRDRI 240
QY 241 PRSPFSGEGFVNFDFYGMFGAQTEDADKTVWTHGNHYHAPNGSLGAMHYESKFRWSE 300
Db 241 PRSPFSGEGFVNFDFYGMFGAQTEDADKTVWTHGNHYHAPNGSLGAMHYESKFRWSE 300
QY 301 GYSDFDRGAVYITFIPIKSWNTAPDKVXKQGP 331
Db 301 GYSDFDRGAVYITFIPIKSWNTAPDKVXKQGP 331

RESULT 3
US-10-112-488-5
; Sequence 5, Application US/10112488
; Publication No. US20030082746A1
; GENERAL INFORMATION:
; APPLICANT: KIKUCHI, Yoshimi
; APPLICANT: DATE, Masayo
; APPLICANT: IMEZAWA, Yukiko
; APPLICANT: YOKOYAMA, Keiichi
; APPLICANT: MATSUI, Hiroshi
; TITLE OF INVENTION: PROCESS FOR PRODUCING TRANSGLUTAMINASE
; FILE REFERENCE: 219286USOCNT
; CURRENT APPLICATION NUMBER: US/10/112,488
; CURRENT FILING DATE: 2002-04-01
; PRIOR APPLICATION NUMBER: PCT/JP00/06780
; PRIOR FILING DATE: 2000-09-29
; PRIOR APPLICATION NUMBER: JP2000-280098
; PRIOR FILING DATE: 2000-06-28
; PRIOR APPLICATION NUMBER: JP11-280098
; PRIOR FILING DATE: 1999-09-30
; NUMBER OF SEQ ID NOS: 70
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 5
; LENGTH: 331
; TYPE: PRT
; ORGANISM: Streptococcus thermophilus
US-10-112-488-5

Query Match 100.0%; Score 1811; DB 15; Length 331;
Best Local Similarity 100.0%; Pred. No. 5, 5e-167;
Matches 331; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DSDRAVTPPAEPLDRLPDPYRPSYGRATVYNNYIRKQOYVSHRDGRKQOMTEBOEWL 60
Db 1 DSDRAVTPPAEPLDRLPDPYRPSYGRATVYNNYIRKQOYVSHRDGRKQOMTEBOEWL 60
QY 61 SYGCVGVTWWSGQYPTNRLAFASFEDEDRFKNELKNGRPRSGETPAEFGGVAKESFDEE 120
Db 61 SYGCVGVTWWSGQYPTNRLAFASFEDEDRFKNELKNGRPRSGETPAEFGGVAKESFDEE 120
QY 121 KGFQARAEVASVNNRALENAHDESAYLNDLKKELANGDNLREDAARSPFSALRNTPSF 180
Db 121 KGFQARAEVASVNNRALENAHDESAYLNDLKKELANGDNLREDAARSPFSALRNTPSF 180
QY 181 KERNGNHDPSRMKAVIYSKHFWSGQDRSSADRRKYGDPDAFRPAPGTGLVMSRDRI 240
Db 181 KERNGNHDPSRMKAVIYSKHFWSGQDRSSADRRKYGDPDAFRPAPGTGLVMSRDRI 240
QY 241 PRSPFSGEGFVNFDFYGMFGAQTEDADKTVWTHGNHYHAPNGSLGAMHYESKFRWSE 300
Db 241 PRSPFSGEGFVNFDFYGMFGAQTEDADKTVWTHGNHYHAPNGSLGAMHYESKFRWSE 300
QY 301 GYSDFDRGAVYITFIPIKSWNTAPDKVXKQGP 331
Db 301 GYSDFDRGAVYITFIPIKSWNTAPDKVXKQGP 331

RESULT 4
US-09-892-864A-2
; Sequence 2, Application US/09892864A
; Patent No. US20020090673A1
; GENERAL INFORMATION:
; APPLICANT: YOKOYAMA, Keiichi
; APPLICANT: ONO, Kunio
; APPLICANT: EJIMA, Daisuke
; TITLE OF INVENTION: PROCESS FOR PRODUCING TRANSGLUTAMINASE
; FILE REFERENCE: 209524USOCNT
; CURRENT APPLICATION NUMBER: US/09/892,864A
; CURRENT FILING DATE: 2001-06-28
; PRIOR APPLICATION NUMBER: PCT/JP99/07250
; PRIOR FILING DATE: 1999-12-24
; PRIOR APPLICATION NUMBER: JP 10-373131

/ PRIOR FILING DATE: 1998-12-28
 / NUMBER OF SEQ ID NOS: 58
 / SOFTWARE: PatentIn version 3.1
 / SEQ ID NO 2
 / LENGTH: 332
 / TYPE: PRT
 / ORGANISM: Artificial Sequence
 / FEATURE:
 / OTHER INFORMATION: Synthetic Peptide
 US-09-892-864A-2

Query Match 100.0%; Score 1811; DB 9; Length 332;
 Best Local Similarity 100.0%; Pred. No. 5.5e-167;
 Matches 331; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DSDDRVTPPAEPLDMPDPYRPSYGRAETVNNYIRKQOYVSHRDRKQOQWTEBQREWL 60
 DB DSDDRVTPPAEPLDMPDPYRPSYGRAETVNNYIRKQOYVSHRDRKQOQWTEBQREWL 61
 QY 61 SYGCVGTWVNSGOYPTNRLAFASFEDEFRKNELKNGRPRSGETRAEFEGRVAKESFDEE 120
 DB 62 SYGCVGTWVNSGOYPTNRLAFASFEDEFRKNELKNGRPRSGETRAEFEGRVAKESFDEE 121
 QY 121 KGFQARAEVAVNNALENADHESAYLDNLKKELANGDALRNEDARSPFYSALENTPSF 180
 DB 122 KGFQARAEVAVNNALENADHESAYLDNLKKELANGDALRNEDARSPFYSALENTPSF 181
 QY 181 KERNGNHDSRKAKVITSKHFMSGQDRSSADKRYGDPDAFPAGTGLVMSRDRNI 240
 DB 182 KERNGNHDSRKAKVITSKHFMSGQDRSSADKRYGDPDAFPAGTGLVMSRDRNI 241
 QY 241 PRSPTSGEGFVNFYGFAGQTEADADKTWTHGNHYHAENSLGAMHYESKFRNMSE 300
 DB 242 PRSPTSGEGFVNFYGFAGQTEADADKTWTHGNHYHAENSLGAMHYESKFRNMSE 301
 QY 301 GYSPFDRGAVITTFIPKSNMTAPDKYQGWMP 331
 DB 302 GYSPFDRGAVITTFIPKSNMTAPDKYQGWMP 332

RESULT 5

US-10-124-429-4
 / Sequence 4, Application US/10124429
 / Publication No. US20020187525A1
 / GENERAL INFORMATION:
 / APPLICANT: TAGUCHI, SEIICHI
 / APPLICANT: MOMOSE, HARUO
 / TITLE OF INVENTION: A METHOD OF PRODUCING MICROBIAL TRANSGLUAMINASE
 / FILE REFERENCE: 220567USOCONT
 / CURRENT APPLICATION NUMBER: US/10/124,429
 / CURRENT FILING DATE: 2002-04-18
 / PRIOR APPLICATION NUMBER: PCT/JP00/07135
 / PRIOR FILING DATE: 2000-10-13
 / PRIOR APPLICATION NUMBER: JP 11-295649
 / PRIOR FILING DATE: 1999-10-18
 / NUMBER OF SEQ ID NOS: 7
 / SOFTWARE: PatentIn version 3.1
 / SEQ ID NO 4
 / LENGTH: 407
 / TYPE: PRT
 / ORGANISM: Streptococcus mobaraense
 US-10-124-429-4

Query Match 100.0%; Score 1811; DB 14; Length 407;
 Best Local Similarity 100.0%; Pred. No. 7.4e-167;
 Matches 331; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DSDDRVTPPAEPLDMPDPYRPSYGRAETVNNYIRKQOYVSHRDRKQOQWTEBQREWL 60
 DB 77 DSDDRVTPPAEPLDMPDPYRPSYGRAETVNNYIRKQOYVSHRDRKQOQWTEBQREWL 136
 QY 61 SYGCVGTWVNSGOYPTNRLAFASFEDEFRKNELKNGRPRSGETRAEFEGRVAKESFDEE 120

DB 137 SYGCVGTWVNSGOYPTNRLAFASFEDEFRKNELKNGRPRSGETRAEFEGRVAKESFDEE 196
 QY 121 KGFQARAEVAVNNALENADHESAYLDNLKKELANGDALRNEDARSPFYSALENTPSF 180
 DB 197 KGFQARAEVAVNNALENADHESAYLDNLKKELANGDALRNEDARSPFYSALENTPSF 256
 QY 181 KERNGNHDSRKAKVITSKHFMSGQDRSSADKRYGDPDAFPAGTGLVMSRDRNI 240
 DB 257 KERNGNHDSRKAKVITSKHFMSGQDRSSADKRYGDPDAFPAGTGLVMSRDRNI 316
 QY 241 PRSPTSGEGFVNFYGFAGQTEADADKTWTHGNHYHAENSLGAMHYESKFRNMSE 300
 DB 317 PRSPTSGEGFVNFYGFAGQTEADADKTWTHGNHYHAENSLGAMHYESKFRNMSE 376
 QY 301 GYSPFDRGAVITTFIPKSNMTAPDKYQGWMP 331
 DB 377 GYSPFDRGAVITTFIPKSNMTAPDKYQGWMP 407

RESULT 6

US-10-112-488-13
 / Sequence 13, Application US/10112488
 / Publication No. US20030082746A1
 / GENERAL INFORMATION:
 / APPLICANT: KIKUCHI, Yoshimi
 / APPLICANT: DATE, Masayo
 / APPLICANT: UMEZAWA, Yukiko
 / APPLICANT: YOKOYAMA, Keiichi
 / APPLICANT: MATSUI, Hiroshi
 / TITLE OF INVENTION: PROCESS FOR PRODUCING TRANSGLUAMINASE
 / FILE REFERENCE: 219286USOCONT
 / CURRENT APPLICATION NUMBER: US/10/112,488
 / CURRENT FILING DATE: 2002-04-01
 / PRIOR APPLICATION NUMBER: PCT/JP00/06780
 / PRIOR FILING DATE: 2000-09-29
 / PRIOR APPLICATION NUMBER: JP2000-280098
 / PRIOR FILING DATE: 2000-06-28
 / PRIOR APPLICATION NUMBER: JP11-280098
 / PRIOR FILING DATE: 1999-09-30
 / NUMBER OF SEQ ID NOS: 70
 / SOFTWARE: PatentIn version 3.1
 / SEQ ID NO 13
 / LENGTH: 407
 / TYPE: PRT
 / ORGANISM: Streptococcus mobaraense
 US-10-112-488-13

Query Match 100.0%; Score 1811; DB 15; Length 407;
 Best Local Similarity 100.0%; Pred. No. 7.4e-167;
 Matches 331; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DSDDRVTPPAEPLDMPDPYRPSYGRAETVNNYIRKQOYVSHRDRKQOQWTEBQREWL 60
 DB 77 DSDDRVTPPAEPLDMPDPYRPSYGRAETVNNYIRKQOYVSHRDRKQOQWTEBQREWL 136
 QY 61 SYGCVGTWVNSGOYPTNRLAFASFEDEFRKNELKNGRPRSGETRAEFEGRVAKESFDEE 120
 DB 137 SYGCVGTWVNSGOYPTNRLAFASFEDEFRKNELKNGRPRSGETRAEFEGRVAKESFDEE 196
 QY 121 KGFQARAEVAVNNALENADHESAYLDNLKKELANGDALRNEDARSPFYSALENTPSF 180
 DB 197 KGFQARAEVAVNNALENADHESAYLDNLKKELANGDALRNEDARSPFYSALENTPSF 256
 QY 181 KERNGNHDSRKAKVITSKHFMSGQDRSSADKRYGDPDAFPAGTGLVMSRDRNI 240
 DB 257 KERNGNHDSRKAKVITSKHFMSGQDRSSADKRYGDPDAFPAGTGLVMSRDRNI 316
 QY 241 PRSPTSGEGFVNFYGFAGQTEADADKTWTHGNHYHAENSLGAMHYESKFRNMSE 300
 DB 317 PRSPTSGEGFVNFYGFAGQTEADADKTWTHGNHYHAENSLGAMHYESKFRNMSE 376
 QY 301 GYSPFDRGAVITTFIPKSNMTAPDKYQGWMP 331

Db 377 GYSDPDRGAYVITFIPKSMWNTAPDKYKQGP 407

RESULT 7

US-10-022-809-2

Sequence 2, Application US/10022809

Publication No. US20030113407A1

GENERAL INFORMATION:

APPLICANT: LIN, Yi-Shin

APPLICANT: LIU, Chang-Hsiueh

APPLICANT: CHU, Wen-Shen

TITLE OF INVENTION: TRANSGUTAMINASE GENE OF STREPTOCOCCUS LADAKANUM AND THE

TITLE OF INVENTION: TRANSGUTAMINASE ENCODED THEREFROM

FILE REFERENCE: U-013779-2

CURRENT APPLICATION NUMBER: US/10/022, 809

CURRENT FILING DATE: 2002-03-29

PRIOR APPLICATION NUMBER: US 10/021, 678

PRIOR FILING DATE: 2001-12-12

NUMBER OF SEQ ID NOS: 3

SEQ ID NO 2

LENGTH: 410

TYPE: PRT

ORGANISM: Streptococcus lactis

US-10-022-809-2

Query Match 95.3%; Score 1726; DB 15; Length 410;

Best Local Similarity 93.4%; Pred. No. 1.3e-158;

Matches 309; Conservative 14; Mismatches 8; Indels 0; Gaps 0;

QY 1 DSDRTTPAEPLDMPDPYPSYGRATYVNNYIRKQOYSHRDGKQMTBEBREML 60

Db 80 DSDRTTPAEPLDMPDPYPSYGRATYVNNYIRKQOYSHRDGKQMTBEBREML 139

QY 61 SYGCVGWVNSGQYPTNRLAFASFDDEFRKNEKNGRPSGSETRAEFEGRVAKESFDEB 120

Db 140 SYGCVGWVNSGQYPTNRLAFASFDDEFRKNEKNGRPSGSETRAEFEGRVAKESFDEB 199

QY 121 KFORAREVAVNNRANALNADHESAYLDNKKELANGNDALRNEDASPFYSALRNTPSF 180

Db 200 KFORAREVAVNNRANALNADHESAYLDNKKELANGNDALRNEDASPFYSALRNTPSF 259

QY 161 KERNGNHDPSSMKAVIYSKHFWSGQDRSSADKRYGDPDAFRPAGTGLVMSDRNI 240

Db 260 KERNGNHDPSSMKAVIYSKHFWSGQDRSSADKRYGDPDAFRPAGTGLVMSDRNI 319

QY 241 PRSPTSPGSEFVNFYGMFGAQTADADKTWTHGNHYHAPNGSLGAMHYESKFPNMSB 300

Db 320 PRSPTSPGSEFVNFYGMFGAQTADADKTWTHGNHYHAPNGSLGAMHYESKFPNMSB 379

QY 301 GYSDPDRGAYVITFIPKSMWNTAPDKYKQGP 331

Db 380 GYSDPDRGAYVITFIPKSMWNTAPDKYKQGP 410

RESULT 8

US-10-112-488-43

Sequence 43, Application US/10112488

Publication No. US20030082746A1

GENERAL INFORMATION:

APPLICANT: KIKUCHI, Yoshimi

APPLICANT: DATE, Masayo

APPLICANT: UMEZAWA, Yukiko

APPLICANT: YOKOYAMA, Keiichi

APPLICANT: NAKSUI, Hiroshi

TITLE OF INVENTION: PROCESS FOR PRODUCING TRANSGUTAMINASE

FILE REFERENCE: 219286USCONT

CURRENT APPLICATION NUMBER: US/10/112, 488

CURRENT FILING DATE: 2002-04-01

PRIOR APPLICATION NUMBER: PCT/JP00/06780

PRIOR FILING DATE: 2000-09-29

PRIOR APPLICATION NUMBER: JP2000-280098

PRIOR FILING DATE: 2000-06-28

PRIOR APPLICATION NUMBER: JP11-280098

QY 2 SDRRTTPAEPLDMPDPYPSYGRATYVNNYIRKQOYSHRDGKQMTBEBREML 61

Db 1 SDRRTTPAEPLDMPDPYPSYGRATYVNNYIRKQOYSHRDGKQMTBEBREML 60

QY 62 YGCVGWVNSGQYPTNRLAFASFDDEFRKNEKNGRPSGSETRAEFEGRVAKESFDEB 121

Db 61 YGCVGWVNSGQYPTNRLAFASFDDEFRKNEKNGRPSGSETRAEFEGRVAKESFDEB 120

QY 122 GFORAREVAVNNRANALNADHESAYLDNKKELANGNDALRNEDASPFYSALRNTPSF 181

Db 121 GFORAREVAVNNRANALNADHESAYLDNKKELANGNDALRNEDASPFYSALRNTPSF 180

QY 182 ERNGNHDPSMKAVIYSKHFWSGQDRSSADKRYGDPDAFRPAGTGLVMSDRNI 241

Db 181 ERNGNHDPSMKAVIYSKHFWSGQDRSSADKRYGDPDAFRPAGTGLVMSDRNI 240

QY 242 RSPSTSPGSEFVNFYGMFGAQTADADKTWTHGNHYHAPNGSLGAMHYESKFPNMSB 301

Db 241 RSPSTSPGSEFVNFYGMFGAQTADADKTWTHGNHYHAPNGSLGAMHYESKFPNMSB 300

QY 302 YSDPDRGAYVITFIPKSMWNTAPDKYKQGP 331

Db 301 YSDPDRGAYVITFIPKSMWNTAPDKYKQGP 330

Query Match 82.1%; Score 1487; DB 15; Length 330;

Best Local Similarity 81.5%; Pred. No. 1.3e-135;

Matches 269; Conservative 26; Mismatches 35; Indels 0; Gaps 0;

RESULT 9

US-10-124-429-2

Sequence 2, Application US/10124429

Publication No. US20020187525A1

GENERAL INFORMATION:

APPLICANT: TAGUCHI, SEIICHI

APPLICANT: MOMOSE, HARUO

TITLE OF INVENTION: A METHOD OF PRODUCING MICROBIAL TRANSGUTAMINASE

FILE REFERENCE: 220567USCONT

CURRENT APPLICATION NUMBER: US/10/124, 429

CURRENT FILING DATE: 2002-04-18

PRIOR APPLICATION NUMBER: PCT/JP00/07135

PRIOR FILING DATE: 2000-10-13

PRIOR APPLICATION NUMBER: JP 11-295649

PRIOR FILING DATE: 1999-10-18

NUMBER OF SEQ ID NOS: 7

SOFTWARE: PatentIn version 3.1

SEQ ID NO 2

LENGTH: 416

TYPE: PRT

ORGANISM: Streptococcus cinamomeum

US-10-124-429-2

Query Match 82.1%; Score 1487; DB 14; Length 416;

Best Local Similarity 81.5%; Pred. No. 1.8e-135;

Matches 269; Conservative 26; Mismatches 35; Indels 0; Gaps 0;

QY 2 SDRRTTPAEPLDMPDPYPSYGRATYVNNYIRKQOYSHRDGKQMTBEBREML 61

Db 87 SDRRTTPAEPLDMPDPYPSYGRATYVNNYIRKQOYSHRDGKQMTBEBREML 146

QY 62 YGCVGWVNSGQYPTNRLAFASFDDEFRKNEKNGRPSGSETRAEFEGRVAKESFDEB 121

Db 147 YGCVGWVNSGQYPTNRLAFASFDDEFRKNEKNGRPSGSETRAEFEGRVAKESFDEB 206

```

QY 122 GFORARVAVSVNNALENADHSAVLDNLKKELANGDALRNEDASPTYSALNTSPK 181
DB 207 GFRARADVAVSVNNALENADHSGTIVNNLKELTNNNDALRLRDSKSNYSALNTSPK 266
QY 182 ERNGNDHPSNRKAVIYSKIFPMGQDRSSADKRYGDPAPAPAGTGLVDMGRDNP 241
DB 267 ERNGNDHPSNRKAVIYSKIFPMGQDRSSADKRYGDPAPAPAGTGLVDMGRDNP 326
QY 242 RSETPSGRGVNFYDYGFGQGTADADKTWTHGNHYHAPNSLGMHYBSKFRNWSG 301
DB 327 RSPARKRGVNFYDYGFGQGTADADKTWTHGNHYHAPNSLGMHYBSKFRNWSG 386
QY 302 YSDPRGAVVITFIPKSNMTAPDKYQGM 331
DB 387 YADPRGAVVITFIPKSNMTAPDKYQGM 416

```

RESULT 10

```

US-10-291-583-92
; Sequence 92, Application US/10291583
; Publication No. US20030138772A1
; GENERAL INFORMATION:
; APPLICANT: Gao, Guangping
; APPLICANT: Wilson, James M.
; APPLICANT: Alviria, Mauricio
; TITLE OF INVENTION: A Method of Detecting and/or Identifying Adeno-Associated Virus
; FILE REFERENCE: UPN-02735USA
; CURRENT APPLICATION NUMBER: US/10/291,583
; CURRENT FILING DATE: 2002-11-12
; PRIOR APPLICATION NUMBER: US 60/350,607
; PRIOR FILING DATE: 2001-11-13
; PRIOR APPLICATION NUMBER: US 60/341,117
; PRIOR FILING DATE: 2001-12-17
; PRIOR APPLICATION NUMBER: US 60/377,066
; PRIOR FILING DATE: 2002-05-01
; PRIOR APPLICATION NUMBER: US 60/386,675
; PRIOR FILING DATE: 2002-06-05
; NUMBER OF SEQ ID NOS: 120
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 92
; LENGTH: 738
; TYPE: PRT
; ORGANISM: capsid protein of AAV serotype, clone 43.1
US-10-291-583-92

```

Query Match 6.8%; Score 122.5; DB 12; Length 738;

Best Local Similarity 23.0%; Pred. No. 0.0055;

Matches 78; Conservative 45; Mismatches 155; Indels 61; Gaps 19;

```

QY 18 DPTPSYGAETVNNYTRKQVYSHDRKQKMTBQREWLSYGVGVTVWNSGQY-- 75
DB 4 DGYLPDM--LENDLSEGRW---WDLKGAAPKPAKQKDDRGJVLDPGYKYLGPFG 58
QY 76 -----PTNRLAFASDEDR-FKNELKNG-RP--RSGETRAFEGRVAKESFDEKQFQRA 126
DB 59 LDKGEFVNADAAALHDKAYDQGLKAGNPIYLRVNHADAFQGRLOE---DTSFGNIG 115
QY 127 REVASVNNRLNENADHSAVLDNLKKELANGDALRNEDASPTYSALNTSPKFRNWSG 186
DB 116 RAYFOAKKRVLEPL-----GLVEBGAKTAPGKKRPVPSRPSDSTGIGKK---- 163
QY 187 NHDPSPKAVIYSKIFPMGQDRSSADKRYGDPAPAPAGTGLVDMGRDNPSPK 246
DB 164 GHOPARKL-----NFGQTSSESVDPPIGPPRPGSGLSGSG--TMAAGGAPMADNN 216
QY 247 PG-----EGFVNDYGVFQAQTEADADKTW--TGNNHY--APNSLGM--HYV 291
DB 217 EGADVSSSGMWHCDSTWLDGVITTSRT-WALPTNNHLKYQJNSGTSGSTNDNTY 275
QY 292 ESKFRWSEGSYDPRGAVVITFIPKSNMTAPDKYQGM 330
DB 276 FGSTPW--GYFDNR--FHCHFSRDMQRL--INNWN 307

```

RESULT 11

```

US-10-291-583-93
; Sequence 93, Application US/10291583
; Publication No. US20030138772A1
; GENERAL INFORMATION:
; APPLICANT: Gao, Guangping
; APPLICANT: Wilson, James M.
; APPLICANT: Alviria, Mauricio
; TITLE OF INVENTION: A Method of Detecting and/or Identifying Adeno-Associated Virus
; FILE REFERENCE: UPN-02735USA
; CURRENT APPLICATION NUMBER: US/10/291,583
; CURRENT FILING DATE: 2002-11-12
; PRIOR APPLICATION NUMBER: US 60/350,607
; PRIOR FILING DATE: 2001-11-13
; PRIOR APPLICATION NUMBER: US 60/341,117
; PRIOR FILING DATE: 2001-12-17
; PRIOR APPLICATION NUMBER: US 60/377,066
; PRIOR FILING DATE: 2002-05-01
; PRIOR APPLICATION NUMBER: US 60/386,675
; PRIOR FILING DATE: 2002-06-05
; NUMBER OF SEQ ID NOS: 120
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 93
; LENGTH: 738
; TYPE: PRT
; ORGANISM: capsid protein of AAV serotype, clone 43.12
US-10-291-583-93

```

Query Match 6.8%; Score 122.5; DB 12; Length 738;

Best Local Similarity 23.0%; Pred. No. 0.0055;

Matches 78; Conservative 45; Mismatches 155; Indels 61; Gaps 19;

```

QY 18 DPTPSYGAETVNNYTRKQVYSHDRKQKMTBQREWLSYGVGVTVWNSGQY-- 75
DB 4 DGYLPDM--LENDLSEGRW---WDLKGAAPKPAKQKDDRGJVLDPGYKYLGPFG 58
QY 76 -----PTNRLAFASDEDR-FKNELKNG-RP--RSGETRAFEGRVAKESFDEKQFQRA 126
DB 59 LDKGEFVNADAAALHDKAYDQGLKAGNPIYLRVNHADAFQGRLOE---DTSFGNIG 115
QY 127 REVASVNNRLNENADHSAVLDNLKKELANGDALRNEDASPTYSALNTSPKFRNWSG 186
DB 116 RAYFOAKKRVLEPL-----GLVEBGAKTAPGKKRPVPSRPSDSTGIGKK---- 163
QY 187 NHDPSPKAVIYSKIFPMGQDRSSADKRYGDPAPAPAGTGLVDMGRDNPSPK 246
DB 164 GHOPARKL-----NFGQTSSESVDPPIGPPRPGSGLSGSG--TMAAGGAPMADNN 216
QY 247 PG-----EGFVNDYGVFQAQTEADADKTW--TGNNHY--APNSLGM--HYV 291
DB 217 EGADVSSSGMWHCDSTWLDGVITTSRT-WALPTNNHLKYQJNSGTSGSTNDNTY 275
QY 292 ESKFRWSEGSYDPRGAVVITFIPKSNMTAPDKYQGM 330
DB 276 FGSTPW--GYFDNR--FHCHFSRDMQRL--INNWN 307

```

RESULT 12

```

US-10-291-583-94
; Sequence 94, Application US/10291583
; Publication No. US20030138772A1
; GENERAL INFORMATION:
; APPLICANT: Gao, Guangping
; APPLICANT: Wilson, James M.
; APPLICANT: Alviria, Mauricio
; TITLE OF INVENTION: A Method of Detecting and/or Identifying Adeno-Associated Virus
; FILE REFERENCE: UPN-02735USA
; CURRENT APPLICATION NUMBER: US/10/291,583
; CURRENT FILING DATE: 2002-11-12

```

PRIOR APPLICATION NUMBER: US 60/350,607
PRIOR FILING DATE: 2001-11-13
PRIOR APPLICATION NUMBER: US 60/341,117
PRIOR FILING DATE: 2001-12-17
PRIOR APPLICATION NUMBER: US 60/377,066
PRIOR FILING DATE: 2002-05-01
PRIOR APPLICATION NUMBER: US 60/386,675
PRIOR FILING DATE: 2002-06-05
NUMBER OF SEQ ID NOS: 120
SOFTWARE: PatentIn version 3.1
SEQ ID NO 94
LENGTH: 738
TYPE: PRT
ORGANISM: capsid protein of AAV serotype, clone 43.5
US-10-291-583-94

Query Match 6.8%; Score 122.5; DB 12; Length 738;
Best Local Similarity 23.0%; Pred. No. 0.0055;
Matches 78; Conservative 45; Mismatches 155; Indels 61; Gaps 19;

QY 18 DPYPSYGAETVNNYIRKQOVYSHRGRKQOMTEBQREWLSCGVGTWVNSGQY-- 75
DB 4 DGYLPDM--LEDNLSEGIHEW---WDLKPGAPKPKANQKQKDDRGVLPEYKYLGFENG 58
QY 76 -----PTNRLAFASDEDR-FKNEIKNG-RP--RSGETRAEFERVAKESTDEKGFQRA 126
DB 59 LDKGEPVNADAAALAEHDKAYDQKAGDNPYLRYNHADAERQERLQE---DTSFGGNIG 115
QY 127 REVASVNNRLENAHDESAVLDNLKKELANGNDALRNEDARSPYSALRNPSPFKERN 186
DB 116 RAVFOAKKRVLEPL-----GLVEBGAKTAPAKKRPEPSPDSSTGIGKK--- 163
QY 187 NHDPSPKAVIYSKHFWSGQDRSSADKRRYGPDAFRPAPGTGLVMSRDRNPRESPTS 246
DB 164 GHQAPARKL-----NFGQGTGDSSEVPDPPIGEPPAGPSGLSAG--TMAAGGAPVADNN 216
QY 247 PG-----EGFVNPYGFPGAGTADADKTW---THGNHYH--APNGSIGAN--HYT 291
DB 217 EGADGVSSSGNWHCDSTWLGDRVITTSRT-WALPTNNHLYKOISNGTSGSTINDNTY 275
QY 292 ESKFRNWSGSDPDGAVYIFPIKSWNTAPDKVKGW 330
DB 276 FGYSSTPW--GYFPDNR--FHCHFSRDMQRL---INNWW 307

RESULT 13

US-10-291-583-2
Sequence 2, Application US/10291583
Publication No. US20030138772A1

GENERAL INFORMATION:
APPLICANT: Gao, Guangping
APPLICANT: Wilson, James M.
APPLICANT: Alvirra, Mauricio
TITLE OF INVENTION: A Method of Detecting and/or Identifying Adeno-Associated Virus
FILE REFERENCE: UPR-02735USA
CURRENT APPLICATION NUMBER: US/10/291,583
PRIOR FILING DATE: 2002-11-12
PRIOR APPLICATION NUMBER: US 60/350,607
PRIOR FILING DATE: 2001-11-13
PRIOR APPLICATION NUMBER: US 60/341,117
PRIOR FILING DATE: 2001-12-17
PRIOR APPLICATION NUMBER: US 60/377,066
PRIOR FILING DATE: 2002-05-01
PRIOR APPLICATION NUMBER: US 60/386,675
PRIOR FILING DATE: 2002-06-05
NUMBER OF SEQ ID NOS: 120
SOFTWARE: PatentIn version 3.1
SEQ ID NO 2
LENGTH: 737
TYPE: PRT
ORGANISM: capsid protein of adeno-associated virus serotype 7
US-10-291-583-2

Query Match 6.7%; Score 122; DB 12; Length 737;
Best Local Similarity 22.8%; Pred. No. 0.0061;
Matches 76; Conservative 45; Mismatches 161; Indels 52; Gaps 16;

QY 18 DPYPSYGAETVNNYIRKQOVYSHRGRKQOMTEBQREWLSCGVGTWVNSGQY-- 75
DB 4 DGYLPDM--LEDNLSEGIHEW---WDLKPGAPKPKANQKQKDDRGVLPEYKYLGFENG 58
QY 76 -----PTNRLAFASDEDR-FKNEIKNG-RP--RSGETRAEFERVAKESTDEKGFQRA 126
DB 59 LDKGEPVNADAAALAEHDKAYDQKAGDNPYLRYNHADAERQERLQE---DTSFGGNIG 115
QY 127 REVASVNNRLENAHDESAVLDNLKKELANGNDALRNEDARSPYSALRNPSPFKERN 186
DB 116 RAVFOAKKRVLEPL-----GLVEBGAKTAPAKKRPEPSPDSSTGIGKK--- 163
QY 187 NHDPSPKAVIYSKHFWSGQDRSSADKRRYGPDAFRPAPGTGLVMS--RDRNP 241
DB 164 GQGPARKRL-----NFGQGTGDSSEVPDPPIGEPPAASVSGTVAAGGAPVADNNNG 218
QY 242 RSPSPGEGFVNDYGVNFGAQTADADKTW---THGNHYHAPNGSIGAHVYESKERN 298
DB 219 ADGVNAGSNWHCDSTWLGDRVITTSRT-WALPTNNHLYKOISNGTSGSTINDNTYFGY 277
QY 299 SE--GYSDPDGAVYIFPIKSWNTAPDKVKGW 330
DB 278 STPWGYFPDNR--FHCHFSRDMQRL---INNWW 306

RESULT 14

US-10-291-583-99
Sequence 99, Application US/10291583
Publication No. US20030138772A1

GENERAL INFORMATION:
APPLICANT: Gao, Guangping
APPLICANT: Wilson, James M.
APPLICANT: Alvirra, Mauricio
TITLE OF INVENTION: A Method of Detecting and/or Identifying Adeno-Associated Virus
FILE REFERENCE: UPR-02735USA
CURRENT APPLICATION NUMBER: US/10/291,583
PRIOR FILING DATE: 2002-11-12
PRIOR APPLICATION NUMBER: US 60/350,607
PRIOR FILING DATE: 2001-11-13
PRIOR APPLICATION NUMBER: US 60/341,117
PRIOR FILING DATE: 2001-12-17
PRIOR APPLICATION NUMBER: US 60/377,066
PRIOR FILING DATE: 2002-05-01
PRIOR APPLICATION NUMBER: US 60/386,675
PRIOR FILING DATE: 2002-06-05
NUMBER OF SEQ ID NOS: 120
SOFTWARE: PatentIn version 3.1
SEQ ID NO 99
LENGTH: 736
TYPE: PRT
ORGANISM: capsid protein of AAV serotype, clone 43.20
US-10-291-583-99

Query Match 6.5%; Score 117; DB 12; Length 736;
Best Local Similarity 22.3%; Pred. No. 0.0197;
Matches 77; Conservative 46; Mismatches 147; Indels 76; Gaps 21;

QY 18 DPYPSYGAETVNNYIRKQOVYSHRGRKQOMTEBQREWLSCGVGTWVNSGQY-- 75
DB 4 DGYLPDM--LEDNLSEGIHEW---WDLKPGAPKPKANQKQKDDRGVLPEYKYLGFENG 58
QY 76 -----PTNRLAFASDEDR-FKNEIKNG-RP--RSGETRAEFERVAKESTDEKGFQRA 126
DB 59 LDKGEPVNADAAALAEHDKAYDQKAGDNPYLRYNHADAERQERLQE---DTSFGGNIG 115
QY 127 REVASVNNRLE--NAHDESAVLDNLKKELANGNDALRNEDARSPYSALRNPSPFKERN 184

Search completed: September 22, 2003, 12:54:23
Job time : 64 secs

Db 116 RAVFOAKRVLPLGLVEBGAKTAPGKKL-----VEQSPQEDSS 156
QY 185 GG---NHDPGRKAVITYSKHPWGGDRSSADARKYGPDAFRAP- GTGLVDMSEDRN 239
Db 157 SGIKGTGQAPAKKL-----NFGQTGDESEVPDPQLGEPPA---APSGLGPMNTAAGCG 208
QY 240 IPRSPSPG-----EGFVNPDYGMFGAQTADADKTV---THGNHYH--APNGSLGA 287
Db 209 APMAONNEGADGVNNSGWNHDSITWLDGRVITTSRT-WALPTYNHLKQISNGTSG 267
QY 288 M---HYESKFRWSEGYSPDRCAGAVITTFPKSMTAPDKVQGW 330
Db 268 STNDNTYFGYSTPW--GYDPNR--FHCHFSRDMQRL---INNHW 306

RESULT 15
US-10-291-583-72
; Sequence 72, Application US/10291583
; Publication No. US20030138772A1
; GENERAL INFORMATION:
; APPLICANT: Gao, Guangping
; APPLICANT: Wilson, James M.
; APPLICANT: Alviria, Mauricio
; TITLE OF INVENTION: A Method of Detecting and/or Identifying Adeno-Associated Virus
; TITLE OF INVENTION: Sequences and Isolating No. US20030138772A1el Sequences Identif
; FILE REFERENCE: UPN-0273USA
; CURRENT APPLICATION NUMBER: US/10/291,583
; CURRENT FILING DATE: 2002-11-12
; PRIOR APPLICATION NUMBER: US 60/350,607
; PRIOR FILING DATE: 2001-11-13
; PRIOR APPLICATION NUMBER: US 60/341,117
; PRIOR FILING DATE: 2001-12-17
; PRIOR APPLICATION NUMBER: US 60/377,066
; PRIOR FILING DATE: 2002-05-01
; PRIOR APPLICATION NUMBER: US 60/386,675
; PRIOR FILING DATE: 2002-06-05
; NUMBER OF SEQ ID NOS: 120
; SOFTWARE: Patent version 3.1
; SEQ ID NO 72
; LENGTH: 737
; TYPE: PRT
; ORGANISM: capsid protein of AAV serotype, clone 3.3bVPI
US-10-291-583-72

Query Match 6.5%; Score 117; DB 12; Length 737;
Best Local Similarity 22.5%; Pred. No. 0.019;
Matches 75; Conservative 45; Mismatches 162; Indels 52; Gaps 16;

QY 18 DYPKPSYGRATVNNYIRKQOVISHRDGRKQOMTEQREMLSYCYGVTVNWSGY-- 75
Db 4 DGYLDPW--LEDNLSGIREW--WDLKRGAPKPRANQOKQDGRGLVPGYKYLGPFG 58
QY 76 -----PTRLAFASFDEDR--FKNELNG--RP--RSGETRAFEGRVAKESPDEKGFORA 126
Db 59 LDKGFPVAAADALAEHDKAYDQQLNAGDNPLYLRNNADAEFQERLQF---DTSPGNUG 115
QY 127 KEVAVYVNRPALENAHDSAYLDNLKKELANGNDAIRNEDARSPYSALANTPSFKERNG 186
Db 116 RAVFOAKRVLPL-----GLVEBGAKTAPAKKRPVEPSPQSPDSSTGIGKK--- 163
QY 187 NHDPGRKAVITYSKHPWGGDRSSADARKYGPDAFRAPAGTGLVDMSEDRN 241
Db 164 GQGPARKL-----NFGQTGDESEVPDPQLGEPPA---APSGLGPMNTAAGCG 218
QY 242 RSPFSPGEGFVNPDYGMFGAQTADADKTV---THGNHYHAPNGSLGAMHYASKFRNW 298
Db 219 ADGVNAGSNWCHDSTWLDGRVITTSRT-WALPTYNHLKQISSETAGSTNNITYFGY 277
QY 299 SE--GYSDPDRCAGAVITTFPKSMTAPDKVQGW 330
Db 278 STPMGFPDPNR--FHCHFSRDMQRL---INNHW 306

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 22, 2003, 12:42:35 ; Search time 19 Seconds

(without alignments)
1675.360 Million cell updates/sec

Title: US-09-884-948-1

Perfect score: 1811

Sequence: 1 DSDDRVTTPAEPLDMPDPY.....ITFIPKSWNTAPDKVKGQWP 331

Scoring table: BIOSUM62

Gapop 10.0 , Gapept 0.5

Searched: 283308 seqs, 96168682 residues

Total number of hits satisfying chosen parameters: 283308

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :
1: p1r1:*
2: p1r2:*
3: p1r3:*
4: p1r4:*

Prod. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match Length	ID	Description
1	1811	100.0	406 2 JC2089	protein-glutamine
2	1802	99.5	332 2 JC2710	protein-glutamine
3	114.5	6.3	1217 2 S52714	sericinB - silkmo
4	106.5	5.9	918 2 B82486	proteinase VCA0223
5	105.5	5.8	1888 2 T14273	zinc finger protei
6	105	5.8	325 2 T83308	hypothetical prote
7	105	5.8	920 2 T52426	dynamn-like prote
8	104	5.8	1032 2 H96619	protein T30E16.17
9	104	5.7	575 1 S03745	beta-amyase (BC 3
10	102	5.6	219 2 B90045	hypothetical prote
11	102	5.6	504 2 T31784	hypothetical prote
12	101.5	5.6	383 2 A8457	acitflavin resista
13	100.5	5.5	926 2 B37271	A-alpha Y 3 protei
14	100	5.5	517 1 A39038	1-calcesonin, norma
15	99	5.5	2273 2 T14677	calcium channel B1
16	99	5.5	2424 2 T4480	calcium channel B1
17	98.5	5.4	1507 2 B47328	natural killer cel
18	98.5	5.4	2924 2 T18378	variant-specific s
19	98	5.4	350 2 S25525	outer membrane por
20	98	5.4	547 1 A56575	puff-specific nucl
21	98	5.4	1403 1 A47328	natural killer cel
22	98	5.4	3938 2 T42761	Bassoon protein -
23	97.5	5.4	625 2 A34615	profilaggrin - rat
24	97	5.4	572 2 S18732	profilaggrin, 64k -
25	97	5.4	591 2 D89783	RGP-containing 11p
26	96.5	5.3	1097 2 T31504	hypothetical prote
27	96.5	5.3	548 2 T28910	hypothetical prote
28	96	5.3	350 2 AF0543	outer membrane por
29	96	5.3	992 2 B86237	protein F14N23.17

30	96	5.3	1344 2 T42637	hypothetical prote
31	96	5.3	1883 2 T13944	chromodomain helic
32	95	5.2	608 2 T32923	hypothetical prote
33	95	5.2	1253 1 A44400	myosin heavy chain
34	95	5.2	1313 2 A48467	myosin heavy chain
35	94.5	5.2	483 2 T19720	hypothetical prote
36	94	5.2	462 2 H64145	hypothetical prote
37	93.5	5.2	396 2 S13251	tropomn T - fruit
38	93.5	5.2	1261 2 G83162	respiratory nitrat
39	93.5	5.2	1722 1 T78879	retinoblastoma bin
40	93.5	5.2	1993 2 AF1450	probable peptidogl
41	93.5	5.2	2441 2 D71623	erythrocyte membra
42	93	5.1	632 2 E75057	peptidase PAB1418
43	92.5	5.1	506 2 G90000	hypothetical prote
44	92.5	5.1	570 1 S50933	myb-related protei
45	92.5	5.1	707 1 S69781	outer membrane pro

ALIGNMENTS

RESULT 1

JC2089
Protein-glutamine gamma-glutamyltransferase (BC 2.3.2.13) precursor - Streptococcus
N/Alternate names: glutamyl-peptide-amine gamma-glutamyltransferase; transglutaminase
C/Species: Streptococcus sp.
C/Date: 14-Jul-1994 #sequence revision 14-Jul-1994 #text change 07-May-1999
C/Accession: JC2089; JC2090; A46730
R/Washizu, K.; Ando, K.; Koike, S.; Hirose, S.; Matsuda, A.; Takai, H.; Motoki, M.;
Biocst. Biotechnol. Biochem. 58, 82-87, 1994
A/Title: Molecular cloning of the gene for microbial transglutaminase from Streptococ
A/Reference number: JC2089; MUID:94162748; PMID:7765334
A/Accession: JC2089
A/Molecule type: DNA
A/Residues: 1-406 <WAS>
A/Experimental source: strain S-8112
R/Takana, S.; Washizu, K.; Ando, K.; Koike, S.; Takeuchi, K.; Matsui, H.; Motoki, M.;
Biocst. Biotechnol. Biochem. 58, 86-92, 1994
A/Title: Chemical synthesis of the gene for microbial transglutaminase from Streptococ
A/Reference number: JC2090; MUID:94162749; PMID:7765335
A/Accession: JC2090
A/Molecule type: DNA
A/Residues: 76-406 <TAK>
R/Kanaji, T.; Ozaki, H.; Takao, T.; Kawajiri, H.; Ide, H.; Motoki, M.; Shimomishi, Y.
J. Biol. Chem. 268, 11565-11572, 1993
A/Title: Primary structure of microbial transglutaminase from Streptococcus sp. st
A/Reference number: A46730; MUID:93280110; PMID:8093353
A/Accession: A46730
A/Status: preliminary
A/Molecule type: protein
A/Residues: 76-406 <KAN>
A/Experimental source: s-8112
A/Note: sequence extracted from NCBI backbone (NCBI:133222)
C/Comment: This enzyme catalyzes an acyl transfer reaction between a gamma-carboxamide
lar crosslinking of certain proteins by gamma-glutamyl-epsilon-lysine side chain bridges
C/superfamily: protein-glutamine gamma-glutamyltransferase
C/keynotes: aminocyclotransferase; calcium; coagulation; heterocyclamer; homodimer
F/1-16/Domain: signal sequence #status predicted <Sig>
F/19-75/Domain: propeptide #status predicted <PRO>
F/76-406/Product: protein-glutamine gamma-glutamyltransferase #status predicted <MAT>

Query Match	100.0%	Score 1811	DB 2	Length 406
Best Local Similarity	100.0%	Pred. No. 2.1e-128		
Matches 331	Conservative	0	Mismatches 0	Indels 0
			Gaps 0	
QY	1	DSDDRVTTPAEPLDMPDPYPSYGAETVNNYIRKQGVYSHRDRKQMTBQREWL	60	
DB	76	DSDDRVTTPAEPLDMPDPYPSYGAETVNNYIRKQGVYSHRDRKQMTBQREWL	135	
QY	61	SYGCVGVTVWNSGVPTNRLAFASFDPRFKNELKNSRPRSGEFTRAFFERVAKESFDEE	120	
DB	136	SYGCVGVTVWNSGVPTNRLAFASFDPRFKNELKNSRPRSGEFTRAFFERVAKESFDEE	195	

QY 121 KGFQARREVASVNNRALLENHDSAYLNDLKKELANGNDALRNEDARSFFYALNRTSPF 180
 DB 196 KGFQARREVASVNNRALLENHDSAYLNDLKKELANGNDALRNEDARSFFYALNRTSPF 255
 QY 181 KERNGNHDPSSMKAVIYSHGFWSGQDRSSADKKYDPPARPPAPGGLVDMRDNI 240
 DB 256 KERNGNHDPSSMKAVIYSHGFWSGQDRSSADKKYDPPARPPAPGGLVDMRDNI 315
 QY 241 PRSPTSPGEGFVNFDMFGAQTADADKTWVTGHNHYHAPNGSLGAMHVYSKFRNWE 300
 DB 316 PRSPTSPGEGFVNFDMFGAQTADADKTWVTGHNHYHAPNGSLGAMHVYSKFRNWE 375
 QY 301 GYSDFRGAAYITFIPIKSNNTAPDKYKQGP 331
 DB 376 GYSDFRGAAYITFIPIKSNNTAPDKYKQGP 406

RESULT 2

UC7310
 protein-glutamine gamma-glutamyltransferase (EC 2.3.2.13) - Escherichia coli
 N:Alternate names: microbial transglutaminase
 C:Species: Escherichia coli
 C:Date: 08-Sep-2000 #sequence_revision 08-Sep-2000 #text_change 17-Nov-2000
 C:Accession: J07310
 R:Yokoyama, K.; Nakamura, N.; Seguro, K.; Kubota, K.
 Biosci. Biotechnol. Biochem. 64, 1263-1270, 2000
 A:Title: Overproduction of microbial transglutaminase in Escherichia coli, in vitro refc
 A:Reference number: J07310
 A:Accession: J07310
 A:Molecule type: DNA
 A:Residues: 1-332 <YOK>
 A:Experimental source: strain JM109
 A:Note: Cross-reference
 C:Comment: This enzyme belongs to a family of enzymes that catalyzes acyl transfer betwe
 in the formation of epsilon-(gamma-glutamyl) lysine cross-linkages. This enzyme is invc
 C:Genetics:
 A:Gene: mtg
 C:Superfamily: Protein-glutamine gamma-glutamyltransferase
 C:Keywords: aminoacyltransferase

Query Match 99.5%; Score 1802; DB 2; Length 332;
 Best Local Similarity 99.7%; Pred. No. 7,5e-128;
 Matches 330; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 DSDDRVTPPAEPLDRMPDYPSPYGAETVANNYIRKQOYVSHRDGRKQMTREOREML 60
 DB 2 DSDDRVTPPAEPLDRMPDYPSPYGAETVANNYIRKQOYVSHRDGRKQMTREOREML 61
 QY 61 SYGCVGVTWVNSGOYPTNRLAFASPEDDFKNEIKNGRPPSGSTRAEFEGRYAKSPDEE 120
 DB 62 SYGCVGVTWVNSGOYPTNRLAFASPEDDFKNEIKNGRPPSGSTRAEFEGRYAKSPDEE 121
 QY 121 KGFQARREVASVNNRALLENHDSAYLNDLKKELANGNDALRNEDARSFFYALNRTSPF 180
 DB 122 KGFQARREVASVNNRALLENHDSAYLNDLKKELANGNDALRNEDARSFFYALNRTSPF 181
 QY 181 KERNGNHDPSSMKAVIYSHGFWSGQDRSSADKKYDPPARPPAPGGLVDMRDNI 240
 DB 182 KERNGNHDPSSMKAVIYSHGFWSGQDRSSADKKYDPPARPPAPGGLVDMRDNI 241
 QY 241 PRSPTSPGEGFVNFDMFGAQTADADKTWVTGHNHYHAPNGSLGAMHVYSKFRNWE 300
 DB 242 PRSPTSPGEGFVNFDMFGAQTADADKTWVTGHNHYHAPNGSLGAMHVYSKFRNWE 301
 QY 301 GYSDFRGAAYITFIPIKSNNTAPDKYKQGP 331
 DB 302 GYSDFRGAAYITFIPIKSNNTAPDKYKQGP 332

RESULT 3
 S52714
 sericinB - silkworm
 C:Species: Bombyx mori (silkworm)

C:Date: 19-May-1995 #sequence_revision 21-Jul-1995 #text_change 09-Sep-1997
 C:Accession: S52714
 R:Garrel, A.A.; Deleage, G.G.; Prudhomme, J.J.
 Submitted to the EMBL Data Library, March 1995
 A:Description: Structure and organisation of the Bombyx mori sericin I gene and of t
 A:Reference number: S52714
 A:Accession: S52714
 A:Status: preliminary
 A:Molecule type: mRNA
 A:Residues: 1-1217 <GAR>
 A:Cross-references: EMBL:Z48802; NID:g75699; PID:g756700

Query Match 6.3%; Score 114.5; DB 2; Length 1217;
 Best Local Similarity 20.4%; Pred. No. 1.8;
 Matches 60; Conservative 47; Mismatches 136; Indels 51; Gaps 11;

QY 33 NYRKQOQVYSHRDGKQKQMTREOR--EMLSGCVGVTWVNSGQ-----YPTNR 79
 DB 124 NYVSDGQAVASDADENRNSAQWQAQNMADSGVGSADRSRRRQANYSDKD 183
 QY 80 LAFASPEDDFKNEIKNGRPPSGSTRAEFEGRYAKSPDEEKGQFQARREVASVNNRALLEN 139
 DB 184 ITAASKDSSADSSRSN-----AYNRDSD-----GSBSAGLSDRSASS 223
 QY 140 AHDSAYLNDLKKELANGNDALRNEDARSFFYALNRTSPFERNGNHDPSS-----RMK 194
 DB 224 SKNINNVHYPTKDSI--GGQAKSPSSSHSGSDAYVNSPPDSYVAGTRDSTSNXXKAS 281
 QY 195 AVIYS--KHFWSGQDRSSADKKYDPPARPPAPGGLVDMRDNI PRSPTSPGEGFV 252
 DB 282 STIYADDQDIRAANDSSSKQKQSSAGISGGP-KGNV--SSKQRYSDNRKRSKSDAYV 338
 QY 253 NFDGVMFG--AQTADADKTWVTGHNHYHAPNGSLGAMHVYSKFRNWESEYSD 304
 DB 333 GRD-----GTAVSNKSEKTSKQSTVNTAQNVSRSDSAASQTSKSDRGYSD 388

RESULT 4

E82486
 proteinase VCA0223 (imported) - Vibrio cholerae (strain N16961 serogroup O1)
 C:Species: Vibrio cholerae
 C:Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 02-Feb-2001
 C:Accession: E82486
 R:Heidelberg, J.F.; Eisen, J.A.; Nelson, W.C.; Clayton, R.A.; Gwinn, M.L.; Dodson, R.
 chardson, D.; Ermolaeva, M.D.; Vamshaveyan, J.; Bass, S.; Qin, H.; Dragoi, I.; Seller
 1, R.R.; Mekalanos, J.J.; Venter, J.C.; Fraser, C.M.
 Nature 406, 477-483, 2000
 A:Title: DNA sequence of both chromosomes of the cholera pathogen Vibrio cholerae.
 A:Reference number: A82035; M01D:20406833; PMID:10952301
 A:Accession: E82486
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-918 <HEI>
 A:Cross-references: GB:AE004362; GB:AE003853; NID:g9657611; PID:NAA96135.1; GSPDB:G
 A:Experimental source: serogroup O1, strain N16961; biotype El Tor
 C:Genetics:
 A:Gene: VCA0223
 A:Map position: 2
 C:Superfamily: Bacillus thuringiensis immune inhibitor A

Query Match 5.9%; Score 106.5; DB 2; Length 918;
 Best Local Similarity 21.2%; Pred. No. 5.1;
 Matches 65; Conservative 52; Mismatches 113; Indels 77; Gaps 16;

QY 42 YSHR-DGRKQMTREOREMLSYGCVGVTWVNSGOYPTNLT-AFASPEDD-REKNEIKNGR 98
 DB 446 YSNNGDLDKRMGRP-----LITPASQQLTFKXMFQTEKDYIARVILNGR 493
 QY 99 PRSGETRA--EFEGRYAKSPDEEKGQFQAR--EVASVNNRALLENHDP-----ESA 145
 DB 494 PIAGNITTMDDPFKSGLVPAISGQSDGWDVDAQDLSAMAGCVFELAFYLLDGLAMBEGL 553
 QY 146 YLDLKKELANGNDALRNEDARSFFYALNRTSPF--ERNGNHDPSSMKAVIYSHGF 203

```

Db      554 YVDDIRLEVDGNQTLIDNAECTSSP-----AFQCFITKNGGFHE-----ANHYX 596
      204 SGQDRSSSS-----AKKRYGDDPDARFAPBTGIVUMSNDRIIPSPISPEGEVFNPDVG 257
      Db      559 LIQMSHNDVDOGLAMTKRFQGLMSPFEGLLIWMYVDESDADVWVK--HPGE-----G 647
      QY      256 WFGQTEADADKTVWTHGNHYHAPNGSLGAMHYVESKFRNWSGYSDFRGAIVTFPIPK 317
      Db      648 WLGV-VIADQNALVWVKTG-----EVAQGRFVQVRAATPSLDDQAP--LKLVTN 692
      QY      318 SWNTAPD 324
      Db      693 DGNLTLED 699

```

RESULT 5

```

T14273
zinc finger protein 106 - mouse
C:Species: Mus musculus (house mouse)
C:Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 20-Sep-1999
C:Accession: T14273
R:Zuberi, A.R.; Christenson, G.J.; Mendoza, L.M.; Shastri, N.; Roopenian, D.C.
Immunol. 9, 687-698, 1998
A:Title: Positional cloning and molecular characterization of an immunodominant cytotoxic
A:Reference number: Z17953; WUID:99060924; PMID:9846490
A:Accession: T14273
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-188 <ZUB>
A:Cross-references: EMBL:AF060246; NID:93372656; P1D:93372657; P1DN:ABD04329.1
C:Genetics:
A:Note: Zfp106

```

Query Match 5.8%; Score 105.5; DB 2; Length 1888;

Best Local Similarity 19.6%; Pred. No. 15; Matches 68; Conservative 37; Mismatches 101; Indels 141; Gaps 17;

```

QY      8 PPAEELRMDPPFRPSYGAETVANNYIRKMQYTHRDGRQOMTEOREMLSYGCVG 67
      Db      109 PPSNSQEVNSDDQPOQRREDRIPIYQDRESYSQPRHRG-----PQGRDW----- 154
      QY      68 TWVNSGQYPTNRLAFASPDDEFKNELEK-GRPR-----SGETRA----- 106
      Db      155 KWEKDG-----FWSTKNSFPISLNSGGPRGSSVHKATGSSSTWFLNHSNGGQ 206
      QY      107 -----EFGRAVAKESFDEE--KGFQRAEIVASVNNRALENHDESAVLDNLKEL 154
      Db      207 WHSNNQWVDMVNYNGTGRNSWSHSEGTGF-----PSWHANN----- 242
      QY      155 ANGDALRNEDASPRYSALRNT-----PSFYERNGNHDPSEM-----X 194
      Db      243 SNGN-----KSSVRSSTNSWNYNGDGFQCGGRNRPNYQMEDMTKMNKKSNK 291
      QY      195 AVITYSKH--FMSGODRSSASADKRYGDDPARPRPGIUVMSRORNPSPSPGPGV 252
      Db      229 PSKTSQERCKKQRODQDKAKYR--SPEEGY-----ASDTPSRGLD 331
      QY      253 NFDYQWFGAQTEADADKTVWTHGNHYHAPNGSLGAMHYVESKFRNWS 299
      Db      333 ERFNF-----EQRESQTKQDVTASKINGKNGTKA-----RDFPRMT 369

```

RESULT 6

```

T38308
hypothetical protein SPAC23H3.15c - fission yeast (Schizosaccharomyces pombe)
C:Species: Schizosaccharomyces pombe
C:Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 20-Oct-2000
C:Accession: T38308; T38381
R:Stelton, V.; Churcher, C.M.; Barrett, B.G.; Rajandream, M.A.; Walsh, S.V.; Wood, V.
submitted to the EMBL Data Library, September 1997
A:Reference number: Z21785
A:Accession: T38308

```

```

A:Molecule type: DNA
A:Residues: 1-325 <EXE>
A:Cross-references: EMBL:Z99163; P1DN:CA816245.1; GSPDB:GN000066; SPDB:SPAC23H3.15c
A:Experimental source: strain 972h-; cosmid c23H3
R:Stelton, V.; Churcher, C.M.; Barrett, B.G.; Rajandream, M.A.; Wood, V.
submitted to the EMBL Data Library, August 1997
A:Reference number: Z21737
A:Accession: T38381
A:Molecule type: DNA
A:Residues: 1-87 <DEPV>
A:Cross-references: EMBL:Z98944; P1DN:CA811598.1; GSPDB:GN000066; SPDB:SPAC23H1.01c
A:Experimental source: strain 972h-; cosmid c23H1
C:Genetics:
A:Gene: SPAC23H3.15c; SPDB:SPAC23H1.01c
A:Map position: 1

```

Query Match 5.8%; Score 105; DB 2; Length 325;

Best Local Similarity 23.0%; Pred. No. 1.7; Matches 64; Conservative 33; Mismatches 85; Indels 96; Gaps 16;

```

QY      71 NSGQYPTNRLAFASPDDEFKFN-----ELANGRPSEGETRAFEGRVAKESFDE 120
      Db      32 SSGAPQNNPFTSTTSEIPENSRRANDMGTDGSDPYAGTSS-----DTK 79
      QY      121 KGF-----QPARVASV--NRALENHDESAVLDNLKELANGNDAL---RNEDARSPF 170
      Db      80 KGFNSVESRKRKQSDVRGDTYSRRHDSYSN--KYSTGNDYSYSGGRVED----- 132
      QY      171 YSLNRTPEFKRNGN--DPSRM-KAVIYGRHWS-----GQDRSSADKPK 216
      Db      133 YST-----SGGSYTTDPSFTDDTASVQSQYNQSRRTTGCGYGEVYSQSYPTDT 182
      QY      217 YGPDFAFRPAPGTCGLVDSRDRNIPSPSPGEGVNPYGMFAQTEADADKTVWTHGN 276
      Db      183 YG-----SRKATPSDITVGGG-----YDSSSSSHIHGSHCTEHRGSS 222
      QY      277 H--YHAPNGSLGAMHYVESKFRNWSBYG--DPRGAY 310
      Db      223 YGNDNTANKTRGAVS-----SAGYSGEGYKGT 251

```

RESULT 7

```

T52426
dynamn-like protein [imported] - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 24-Oct-2000 #sequence_revision 24-Oct-2000 #text_change 24-Oct-2000
C:Accession: T52426
R:Kato, A.; Suzuki, M.; Kuwahara, A.; Ooe, H.; Higano-Inaba, K.; Komeda, Y.
Gene 239, 309-316, 1999
A:Title: Isolation and analysis of cDNA within a 300 kb Arabidopsis thaliana genomic reg
A:Reference number: Z25171
A:Accession: T52426
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-920 <KAT>
A:Cross-references: EMBL:AB028467; P1DN:BA88111.1
A:Experimental source: cultivar Columbia
C:Genetics:
A:Gene: CFI
A:Map position: 1

```

Query Match 5.8%; Score 105; DB 2; Length 920;

Best Local Similarity 24.6%; Pred. No. 6.6; Matches 59; Conservative 27; Mismatches 104; Indels 50; Gaps 11;

```

QY      26 RAETVV--NNYRKMQYVSHRDG-----RKQOMTEQREWLSTYG 63
      Db      684 KASMDVKNEMWINKQYIQARFGQVSSAMQSLSEBSLDKMKRKVDDEBELRMSSGE 743
      QY      64 CVGYTWNSGQYPTN-----RLAFASPDDEFKNELEKGRPPSGETRAEF--EGRVAYE 115
      Db      744 VRGYEAVLNSLAANVPAAVLVCVERKEDMLNCLVYSISAIQNERISLLEQDQNVKR 803

```

Qy 116 SPDEKGFQARREVASVMALENADH-----SAYLDNKKELA---NGNDALNEDAR 167
 Db 804 RRDR---YQKQSLSLTQLTQTL-SHDNRMAAASWSDNGTSSSRTNGSSS-GEDEMM 857
 Qy 168 SPFYALNTPSPFERNGNHDPSRKAVIYSKHPWGGQDSSSADKKRYGPDAPFRPAP 227
 Db 858 NAFNMAASGPDLSLKRYSGGH--SRR---YSDPAQNGEDSSGSGSSSRRTTPNLPAP 911

RESULT 8
 H96619
 Protein T30E16.17 [imported] - Arabidopsis thaliana
 C/Species: Arabidopsis thaliana (mouse-ear cress)
 C/Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 31-Mar-2001
 C/Accession: H96619
 R/Theologos, A.; Ecker, J.R.; Palm, C.D.; Federpiet, N.A.; Kaul, S.; White, O.; Alonso, Chn, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.;ansen, N.F.; Hughes, B.; Hultzer, L.
 Nature 408, 816-820, 2000
 A/Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Llorca, J.S.; Maltli, R.; Marziani, Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
 A/Authors: Salzman, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon, Ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Ventner, J.C.; Davis, R.W.
 A/Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
 A/Reference number: A86141; MUID:21016719; PMID:1130712
 A/Accession: H96619
 A/Status: preliminary
 A/Molecule type: DNA
 A/Residues: 1-1092 <STO>
 A/Cross-references: GB:AE005173; NID:g8778745; PIDN:AAF79753.1; GSPDB:GN00141
 C/Genetics:
 A:Gene: T30E16.17
 A:Map position: 1

Query Match 5.8%; Score 105; DB 2; Length 1092;
 Best Local Similarity 24.6%; Pred. No. 8.3; Mismatches 104; Indels 50; Gaps 11;
 Matches 59; Conservative 27;
 Qy 26 RAETV--NNYIRKQGVYSHRDG-----RQQTEBQREWLSTYG 63
 Db 856 KAESWVDKNEWINKLQVLCARAGQVGSASMSQSLSEGSLDKVRKFPVDEBELKWSQOE 915
 Qy 64 CYGVTVNWSGQYPTN-----RLAFSPEDRKYELKNGPRSGEPRAP--EORVAKK 115
 Db 916 VGVYEAVLNSLAANVKAVALLOQVEKSKEDMLQVLSISALGNRIRISLIQEDQNTKR 975
 Qy 116 SPDEKGFQARREVASVMALENADH-----SAYLDNKKELA---NGNDALNEDAR 167
 Db 976 RRDR---YQKQSLSLTQLTQTL-SHDNRMAAASWSDNGTSSSRTNGSSS-GEDEMM 1029
 Qy 168 SPFYALNTPSPFERNGNHDPSRKAVIYSKHPWGGQDSSSADKKRYGPDAPFRPAP 227
 Db 1030 NAFNMAASGPDLSLKRYSGGH--SRR---YSDPAQNGEDSSGSGSSSRRTTPNLPAP 1083

RESULT 9
 S03745
 beta-amylase (EC 3.2.1.2) precursor - Bacillus circulans
 C/Species: Bacillus circulans
 C/Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 10-Sep-1999
 C/Accession: S03745
 R/Sigens, K.W.
 Mol. Microbiol. 1, 86-91, 1987
 A/Title: Molecular cloning and characterization of the beta-amylase gene from Bacillus C
 A/Reference number: S03745; MUID:88260890; PMID:2455212
 A/Accession: S03745
 A/Status: not compared with conceptual translation
 A:Molecule type: DNA
 A/Residues: 1-575 <SIG1>
 A/Cross-references: GB:Y00523; NID:g39407; PIDN:CAA68578.1; PID:g39408
 C/Function:
 A/Description: catalyzes the hydrolysis of 1,4-glycosidic linkages of starch, removing m

C/Superfamily: beta-amylase
 C/Keywords: glycosidase; hydrolase; polysaccharide degradation
 F:1-36/Domains: signal sequence #status predicted <SIG>
 F:37-575/Product: beta-amylase #status predicted <MAT>

Query Match 5.7%; Score 104; DB 1; Length 575;
 Best Local Similarity 22.9%; Pred. No. 4.3; Mismatches 103; Indels 116; Gaps 19;
 Matches 76; Conservative 37;
 Qy 24 YGRAETVNNYIRKQGVYSHRDGQKQMTBQREWLSTYGCVYV-----NSG 73
 Db 80 WQYESAGDQF--DMSYTYTYADTVKQ-----AGLKNVPIITHRGANVG 124
 Qy 74 Q-----YPTNRLAFSPEDRKYELKNGPRSG-----ETRAFEGRVAKSPDEKGFQ 124
 Db 125 DDQNTPLPSMLWKGADAMQKDE-----SGVNNLSLSPFNSGQKQ-YDE----- 171
 Qy 125 RAREVASVMALENADHSAVLDNKKELANGNDALRNEDASPPYSALNTPSPFERN 184
 Db 172 ---LYASF-----KQNSAYKDWIPIKYLSGSP-----SGELRYPSYYPAA 209
 Qy 185 GGNHDPSPMRKAVIYSKHPWGGQDSSSADKKRYGPDAPAPAGTGLVMSRDRNIPSP 244
 Db 210 GMSY-PARGKFGYVTE---TASAPRTMTTKYGSIDKIRANGVILTSMSQ-----ISP 260
 Qy 245 TSPGSGF---VNPDIYQ-----WF-----GAQTEADADKXTWTW-----GNH 277
 Db 261 PTDSDGFYTGQYNYITVGHDFLSWYQVLNHLGVIAAANKVDFVFCVRIAGAKISGIH 320
 Qy 278 YHAPNGSL--GANEVSESKFRMSGSDPFR 307
 Db 321 WQNNPSPKPSHAEH-----AGGYVDYNR 343

RESULT 10
 B90045
 hypothetical protein [imported] - Staphylococcus aureus (strain N315)
 C/Species: Staphylococcus aureus
 C/Date: 10-May-2001 #sequence_revision 10-May-2001 #text_change 22-Oct-2001
 C/Accession: B90045
 R/Kiroda, M.; Ohta, T.; Uchiyama, I.; Baba, T.; Yuzawa, H.; Kobayashi, I.; Cui, L.;ma, A.; Mizutani-Uji, Y.; Kobayashi, N.; Szwano, T.; Inoue, R.; Kaito, C.; Sekimizu, C.; Shiba, T.; Hattori, M.; Ogasawara, N.; Hayashi, H.; Hiratake, K.
 Lancet 357, 1225-1240, 2001
 A/Title: Whole genome sequencing of methicillin-resistant Staphylococcus aureus.
 A/Reference number: A89758; MUID:21311952; PMID:11418146
 A/Accession: B90045
 A/Status: preliminary
 A:Molecule type: DNA
 A/Residues: 1-219 <KOR>
 A/Cross-references: GB:BA000018; PID:g13702382; PIDN:BA843523.1; GSPDB:GN00149
 A/Experimental source: strain N315
 C/Genetics:
 A:Gene: SA2221

Query Match 5.6%; Score 102; DB 2; Length 219;
 Best Local Similarity 20.8%; Pred. No. 1.7; Mismatches 46; Conservative 42; Indels 60; Gaps 11;
 Matches 46;
 Qy 18 DYPSPY-----GRAETVNNYIRK-----QQVYSHRDGKQKQMTBQRE-- 58
 Db 5 DKIRDSIOYNQONHRRQSEDSYRQYAKGDEBEHPERYNGDYREQITLSEBEKKS 64
 Qy 59 -----WLSYGVGVTVNWSGQYPTNRLAFSPEDRKYELKNGPRSGETRAFEGRV- 112
 Db 65 RSKSKWL-YIIIALIILIVAFVTR--ALANDSDKYSNDPKYSQYKKQYENG-DOQIN 120
 Qy 113 -----AKSPDEKGFQARREVASVMALENADHSA-----YLDNKKELANG 157
 Db 121 QQVDNAKEMI---KNQKXTDDITIKNLQMOIDNKKQGRONKADSKLTQFYQDQINK-LTEA 176
 Qy 158 NDALRNEDASPPYSALNTPS-----PERRNGG 186

Db 177 NNALKNASQKIESMLNDINTKEDSIKSLBSPFDNDNG 217

RESULT 11

T31784
Hypothetical protein F13H6.4 - *Caenorhabditis elegans*

C/Species: *Caenorhabditis elegans*

C/Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 31-Jan-2000

C/Accession: T31784

R/Jones, K.; Kohlmann, P.

submitted to the EMBL Data Library, July 1997

A/Description: The sequence of C. elegans cosmid F13H6.

A/Reference number: Z21085

A/Accession: T31784

A/Status: preliminary; translated from GB/EMBL/DDAJ

A/Molecule type: DNA

A/Residues: 1-504 <JUN>

A/Cross-References: EMBL:AF016437; PIRN:AA55885.1; GSPDB:GN00023; CESP:F13H6.4

A/Experimental source: strain Bristol N2; clone F13H6

C/Genetics:

A/Gene: CESP:F13H6.4

A/Map position: 5

A/Intons: 39/1; 63/3; 92/3; 123/2; 205/1; 264/3; 393/1; 436/2

C/Superfamily: cholinesterase; cholinesterase homology

Query Match 5.6%; Score 102; DB 2; Length 504;

Best Local Similarity 20.5%; Pred. No. 5.1;

Matches 69; Conservative 44; Mismatches 144; Indels 80; Gaps 13;

Db 28 ETVNNYIRKQOVYSHRDGRKQMTBQREMSYSG---CVGTWVNSGQYPTNNLARA 83

149 DSVGNNGFGLMDQTLA-----LKNVQKHSSFGADNCTVFGSAGASATDLSIS 200

84 SPEDPRKDELKNGRPSGRTAEPEGRVAK-----ESFDEKFGQARAVASV---NM 134

201 PHSRDLRFRI-----PISGAYCEPRLRTSKSQAKFRFPALFKFTGSGSTLILEYKX 256

135 RALNNAHDSAYLDNKKELANGDARNEDAR-----SPFYSAIRNTPSFKEKNGNHD 189

257 QSSFTLSD--LRKAPKKQMTGVDEYGVIAAMNPEFSPADAGLAFPKGVYGNTPAN 314

190 PSRKAVIYSKHFWSGODRSSGADRRK-----YGDPAFRPAPGVLVMSRDRNTPRSP 244

315 PEBMHKLFYKXY-EEVDKSDDSAMKRLCEAFG----- 348

245 TSPGEGVNDYGMF-GAQTEDADKTVWTHG-NHYAFNGSL-----GAMHYEERK 294

349 -----LGFNLGVFQSAKSAKXGNDVFLYSEFYHSDGFCGMKDLDPFASNGHELR 401

295 FRWSEGVSPDRGAVVITFLPKSNTPADPKVKQMP 331

402 YLLGEGFYSKFDATKEELEVEKTYTLFNFAYKMP 438

Db

RESULT 12

AH3457

acetylflavin resistance protein E [imported] - *Brucella melitensis* (strain 16M)

C/Species: *Brucella melitensis*

C/Date: 01-Feb-2002 #sequence_revision 01-Feb-2002 #text_change 01-Feb-2002

C/Accession: AH3457

R/DelVecchio, V.G.; Kapratli, V.; Redkar, R.U.; Patra, G.; Mujir, C.; Los, T.; Ivanova,

.; Maruz, M.; Golsman, E.; Selkov, E.; Elizer, P.H.; Hastings, S.; O'Callaghan, D.; Letess

Proc. Natl. Acad. Sci. U.S.A. 99, 443-448, 2002

A/Title: The genome sequence of the facultative intracellular pathogen *Brucella melitensis*

A/Reference number: AD3252; PMID:11756688

A/Accession: AH3457

A/Status: preliminary

A/Molecule type: DNA

A/Residues: 1-383 <KUR>

A/Cross-References: GB:AE008917; PIRN:AA552827.1; PID:g7993667; GSPDB:GN00190

A/Experimental source: strain 16M

C/Genetics:

A/Gene: EME11646

A/Map position: 1

Query Match 5.6%; Score 101.5; DB 2; Length 383;

Best Local Similarity 26.1%; Pred. No. 3.9;

Matches 29; Conservative 21; Mismatches 32; Indels 29; Gaps 3;

Db 63 GCYGTWVNSGQYPTNNLAFASFE-----DRFKNELKNGRPSGRTAEPEGR 111

189 GIVGILPVMAGNYVTAQTSIARLDDSTYLDIWIWPERFAQIVQVPLASTA-FGE 247

112 VAVSEFDEKGFQARAVASVNNALNNAHDSAYLDNKKELANGDARN 162

248 IYKGRIN-----AVDNMLDEASPTLHVAFVNAEDRLR 281

Db

RESULT 13

B37271

A1-alpha Y 3 protein - bracket fungus (*Schizophyllum commune*)

C/Species: *Schizophyllum commune*

C/Date: 07-Feb-1992 #sequence_revision 07-Feb-1992 #text_change 05-Dec-1997

C/Accession: B37271

R/Ullrich, R.C.

submitted to the Protein Sequence Database, October 1991

A/Reference number: A37271

A/Accession: B37271

A/Status: preliminary; not compared with conceptual translation

A/Molecule type: DNA

A/Residues: 1-926

C/Superfamily: unassigned homeobox proteins; homeobox homology

C/Keywords: DNA binding; homeobox; nucleus; transcription regulation

F:148-204/domain: homeobox homology <HGX>

Query Match 5.5%; Score 100.5; DB 2; Length 926;

Best Local Similarity 20.5%; Pred. No. 15;

Matches 71; Conservative 36; Mismatches 122; Indels 117; Gaps 15;

Db 9 PAPPDRPDPYRPSYGAETVNNYIRKQOVYSHRDGRKQMTBQREMSYSGVGT 68

601 PAPPYIR-FDDFAPFVALAKRKARRRKRKKQAKRKAR----- 640

69 WNVNSGQYPTNNLAFASFEDEPRKDELKNGRPSGRTAEPEGRVAKSPDEKGFQAR 126

641 -----EKRARKEKQAKKDRKQKRGGLPR-SPEITLDS-----SRAS 678

129 VASVNNRALNNAHDSAYLDNKKELANGDARNEDARSPFYSAIRNTPSFKEKNGN 188

679 VTSDAATSRKSRTRKRPDSASVAS-----ARTP-----SLSTSS--RRSGTS 724

189 DRS--RMK---AVIYSKHFWSGODRESSND---KRKIGPDA----- 222

725 MPATPRMNESSLPVASDNFVLTGDKVITWTPELMAQLGEBDASGLDEPNQSGFSDML 784

223 -FRPAPGTLVMSRDRNIPRSPTPSGEGVNFYGMFG--AQTEDADKTVWTHGNHYH 279

785 IFSSNDGALGMDTNDYMP-----LGDLSPTQLSFDMMNTSSMDLS 828

Db

280 A-ENGSLGAMHYEERKFNWSEGVSPDRGAVVITFLPKSNTPADPKVKQMP 324

829 TOPAASFDSSSETSSMDFNW-----LTPQCANIAPD 859

Db

RESULT 14

A39038

1-caldesmon, nonmuscle - chicken

N/Alternate names: caldesmon, nonmuscle

C/Species: *Gallus gallus* (chicken)

C/Date: 31-Jul-1991 #sequence_revision 27-Jun-1994 #text_change 22-Jun-1999

C/Accession: A39038

R/Hayashi, K.; Fujio, Y.; Kato, I.; Sobue, K.

J. Biol. Chem. 266, 355-361, 1991

A/Title: Structural and functional relationships between h- and l-caloesmons.

A/Reference number: A39038; MUID:91093148; PMID:1824698

A/Accession: A39038

A:Molecule type: mRNA
A:Residues: 1-517 <HAV>
A:Cross-references: GB:M60620; GB:M38015; NID:g212242; PIDN:AAA48936.1; PID:g212243
A:Experimental source: brain
C:Comment: The binding of caldesmon to F-actin is modulated by calcium and calmodulin.
C:Comment: Two calmodulin molecules can bind to nonoverlapping domains of each caldesmon
C:Superfamily: caldesmon
C:Keywords: actin binding; calmodulin binding; phosphoprotein
P:342,427/Binding site: phosphate (Ser) (covalent) #status predicted
P:433,456/Binding site: phosphate (Thr) (covalent) #status predicted

Query Match

5.5%; Score 100; DB 1; Length 517;
Best Local Similarity 18.7%; Pred. No. 7.4;

Matches 60; Conservative 41; Mismatches 116; Indels 104; Gaps 12;

```
QY 28 ETVVNNYIR-KMOQYVHNRDGRKQKMTTEOREMWSYGVGVTVWNSGOVPTNRILAFASFD 66
DB 151 ETVVTSYQRRNNRQ-DGBEGRKKEKQSEEEK-----PREVPT-----E 188
QY 87 EDRFK-NELKNGRPSGRTAEFEGRVAKESPDEEKGFORAREVASYMNRALENA----- 140
DB 189 ENGVKNNKVKKEAPKE-BMKSVMD--RKRGVPEQKQNGEREITTPKLSKSTENAFGRSN 244
QY 141 -----HDSAYLDNLKKELANGDALLRNEDARSPFYSALRNTPSFXE 182
DB 245 LKGAANAERKSEKLEKQEQEAVELDLKRRERKRKILEREQKKQKEFAEKIRESEE 304
QY 183 RNGNHDPSRMKAVIYSKHPWSGQDRSSSADK-----RKYDDPD 221
DB 305 KKMKEKIEERRALEAEKROKVPEDGVSEKKPFKCFSPKGSLSKTEERAPFLNKAQKS 364
QY 222 AFRPAPGTGLVDMRDR-----NTPRSPTS-----PGEGVNPDYGMFGAQT 263
DB 365 GMPKPAHTTAVWSKIDPSRLDQYTSVAVGNKAKAPAPASDLVPAEGVRNI----- 415
QY 264 EADADKTVTWTHGNHHPNGS 284
DB 416 ----KSMWEKGNVSSPGT 431
```

RESULT 15

146477

calcium channel, Bi-1 - rabbit

C:Species: Oryctolagus cuniculus (domestic rabbit)

C:Date: 16-Aug-1996 #sequence_revision 16-Aug-1996 #text_change 17-Nov-2000

C:Accession: 146477; 146478

R:Mori, Y.; Friedlich, T.; Kim, M.S.; Mikami, A.; Nakai, J.; Ruth, P.; Bosse, E.; Hofman

Nature 350, 398-402, 1991

A:Title: Primary structure and functional expression from complementary DNA of a brain c

A:Reference number: 146477; MUID:91187110; PMID:1849233

A:Accession: 146477

A>Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: mRNA

A:Residues: 1-2273 <MOR>

A:Cross-references: EMBL:X57476; NID:g1522; PIDN:CAA40714.1; PID:g1523

A:Accession: 146478

A>Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: mRNA

A:Residues: 1-1856, 'H', 1858, 'K', 1860-1862, 'SL', 1865-1866, 'VIS', 1870-1876, 'K', 1878-1879, 'A:Cross-references: EMBL:X57688; NID:g1524; PIDN:CAA40871.1; PID:g1525

C:Superfamily: voltage-dependent calcium channel protein alpha-1 chain

Query Match

5.5%; Score 99; DB 2; Length 2273;
Best Local Similarity 26.5%; Pred. No. 60;

Matches 43; Conservative 18; Mismatches 61; Indels 40; Gaps 7;

```
QY 99 PRSGTAEFEGRVAKESPDEEKGFORAREVASYMNRALENAHDSEAYLDNLKKELANGN 158
DB 828 PQENNNNTNKSRYAEPFTVDRLGQRAEDFLKQARHDPADPSAH----- 875
QY 159 DALRNEDARSPFYSA-----LRNTPSFKE-----RNGNHDPSRMKAVIYSKHPWSGQDR 208
DB 876 -AAAGLDARRPWAGSGAEILSRGPGYGRSDHQAREGGLEPPG-----FWEGE-- 922
```

QY 209 SSSADKRYGDDPAPRPAPDTGLVDMRDR-IPRSPTSPGE 249
DB 923 ---AERGAQDPFR-RHARQGVGSGSGSRSGSPRTETADGE 960

Search completed: September 22, 2003, 12:45:29
Job time : 23 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 22, 2003, 12:41:14; Search time 14 Seconds

(without alignments)
1111,846 Million cell updates/sec

Title: US-09-884-948-1

Perfect score: 1811

Sequence: 1 DSDDRVTPPAEDPDRMPDXY.....ITPDKSWNPADPKXKQGP 331

Scoring table: ELOSUM62

Gapop 10.0, Gapext 0.5

Searched: 127863 seqs, 47026705 residues

Total number of hits satisfying chosen parameters: 127863

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

SUMMARIES

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

Result No.	Score	Query Match Length	ID	Description
1	1811	100.0	331	1 TGL STRESS
2	108	6.0	1382	1 IP3A HUMAN
3	104	5.7	573	1 AMTB_BACCI
4	100.5	5.5	926	1 MAY3_SCHCO
5	100	5.5	744	1 JPH3_MOUSE
6	99.5	5.5	943	1 ARS2_MOUSE
7	99	5.5	2424	1 CCA4_RABIT
8	98.5	5.4	1453	1 NKCR_MOUSE
9	98	5.4	350	1 PHOE_SALTY
10	98	5.4	547	1 EX42_DROME
11	98	5.4	1462	1 NKCR_HUMAN
12	96	5.3	350	1 PHOE_SALTY
13	96	5.3	497	1 AMYB_STRCA
14	96	5.3	536	1 SKIP_HUMAN
15	96	5.3	1344	1 IP3A_MOUSE
16	95	5.3	2468	1 MAPB_HUMAN
17	95	5.2	572	1 LMND1_HUMAN
18	94	5.2	462	1 YFHD_HAEN
19	93.5	5.2	396	1 TRT_DROME
20	93.5	5.2	1722	1 RBB2_HUMAN
21	93	5.1	482	1 UZR2_HUMAN
22	92.5	5.1	633	1 MTH_TERTH
23	92.5	5.1	790	1 KTF3_MOUSE
24	92	5.1	660	1 JPH1_MOUSE
25	92	5.1	1101	1 DIAD_HUMAN
26	92	5.1	2459	1 MAPB_RAT
27	91.5	5.1	381	1 NCAP_CVCAI
28	91.5	5.1	383	1 OMS2_SALTY
29	91	5.0	350	1 RGA1_RABIT
30	91	5.0	535	1 YGHI_CABBL
31	91	5.0	636	1 GYR3_THEMA
32	91	5.0	1536	1 SIN3_YEAST
33	90.5	5.0	351	1 PHOE_KLEPN

34	90.5	5.0	367	1 OMPC_ECOLI
35	90.5	5.0	632	1 YGA9_SCHPO
36	90.5	5.0	980	1 BOB1_YEAST
37	90.5	5.0	1131	1 MOG1_CABBL
38	90	5.0	351	1 PHOE_CITFR
39	90	5.0	611	1 IF4B_HUMAN
40	89.5	4.9	537	1 ARP_ELARA
41	89.5	4.9	935	1 KINH_SYNER
42	89.5	4.9	3421	1 TEGU_HYVER
43	89	4.9	559	1 QUA1_HUMAN
44	89	4.9	977	1 DLP1_HUMAN
45	88.5	4.9	532	1 CCB1_DROAC

ALIGNMENTS

[Handwritten signature]

RESULT 1	TGL STRESS	STANDARD	PRT	331 AA
AC	P81453			
DT	15-DEC-1998 (Rel. 37, Created)			
DT	15-DEC-1998 (Rel. 37, Last sequence update)			
DT	15-SEP-2003 (Rel. 42, Last annotation update)			
DE	Protein-glutamine gamma-glutamyltransferase (EC 2.3.2.13)			
DE	(Transglutaminase) (Tgase)			
OS	Streptococcus sp. (strain s-8112)			
OC	Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;			
OC	Streptomyces; Streptomyces; Streptomyces			
OX	NCBI_TaxID=86037			
RN	[1]			
RP	SEQUENCE			
RX	MEDLINE=93280110; Pubmed=8099353			
RA	Kenji T., Ozaki H., Takao T., Kawajiri H., Ide H., Motoki M.,			
RA	Shimomishi Y.,			
RT	Primary structure of microbial transglutaminase from			
RT	Streptococcus sp. strain s-8112			
RL	J. Biol. Chem. 268:11565-11572(1993)			
CC	- FUNCTION: Catalyzes the cross-linking of proteins and the			
CC	configuration of polyamines to proteins.			
CC	- CATALYTIC ACTIVITY: Protein glutamine + L-lysine = protein N(5)-			
CC	alkylglutamine + NH(3)			
CC	- MASS SPECTROMETRY: MW=37869.2, MW_ERR=8.8, METHOD=Electrospray,			
CC	- BIOTECHNOLOGY: Sold under the name Activa TG by Ajinomoto. It has			
CC	the ability to crosslink protein molecules present in food without			
CC	the use of salt or binders. Used to improve some of the physical			
CC	properties such as firmness, elasticity and moisture retention of			
CC	food such as meat, poultry and seafood.			
DR	PDB; 1J14; 27-AUG-02.			
KW	Transferase; Acyltransferase; 3D-structure.			
FT	ACT SITE 64			
FT	POTENTIAL 64			
SO	SEQUENCE 331 AA; 37862 MW; 59923636362C1FA CRC64;			
Query Match	100.0%; Score 1811; DB 1; Length 331;			
Best Local Similarity	Pred. No. 1.5e-129; Mismatches 0; Gaps 0;			
Matches 331; Conservative 0; Mismatches 0; Indels 0; Gaps 0;				
QY	1 DSDDRVTPPAEDPDRMPDXY.....ITPDKSWNPADPKXKQGP 60			
DB	1 DSDDRVTPPAEDPDRMPDXY.....ITPDKSWNPADPKXKQGP 60			
QY	61 SVGCVGVWVNSGVYTNELAFSPDEPRFNLKNGRSGSETAFEEGVAAESFDES 120			
DB	61 SVGCVGVWVNSGVYTNELAFSPDEPRFNLKNGRSGSETAFEEGVAAESFDES 120			
QY	121 KQFQAREVAVSVNRLAENADESAVLDNLKELANGNDALNEDARSPPYSALNTPSF 180			
DB	121 KQFQAREVAVSVNRLAENADESAVLDNLKELANGNDALNEDARSPPYSALNTPSF 180			
QY	181 KERNGCNHDSRKAVITYSKHFWSGQDRSSADKRYGPDPAFRAPPTGIVMSRDRNI 240			
DB	181 KERNGCNHDSRKAVITYSKHFWSGQDRSSADKRYGPDPAFRAPPTGIVMSRDRNI 240			


```

QY 241 PRSPTSGEGFVNFEDYWGFGAQTADAKTWTGHNHYHAPNGSIGAMHYESKPERNMS 300
DB 241 PRSPTSGEGFVNFEDYWGFGAQTADAKTWTGHNHYHAPNGSIGAMHYESKPERNMS 300
QY 301 GYSPDFRGAVITTFPKSWNTAPDPKXQGWP 331
DB 301 GYSPDFRGAVITTFPKSWNTAPDPKXQGWP 331

RESULT 2
IP3A_HUMAN STANDARD; PRT; 1382 AA.
ID 1F3A_HUMAN
AC Q14152; 000653;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Eukaryotic translation initiation factor 3 subunit 10 (eIF-3 theta)
DE (eif3 p167) (eif3 p160) (eif3 p185) (eif3a).
GN EIF3S10 OR KIAA0139.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Bone marrow;
RX MEDLINE=96127530; PubMed=8590280;
RA Nagase T., Seki N., Tanaka A., Ishikawa K.-I., Nomura N.;
RT "Prediction of the coding sequences of unidentified human genes. IV.
RT The coding sequences of 40 new genes (KIAA0121-KIAA0160) deduced by
RT analysis of cDNA clones from human cell line KG-1.";
RL DNA Res. 2:167-174(1995).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=97294683; PubMed=9150439;
RA Scholler J.K., Kanner S.B.;
RT "The human p167 gene encodes a unique structural protein that contains
RT centrosomal A homology and associates with a multicomponent complex.";
RL DNA Cell Biol. 16:515-531(1997).
RN [3]
RP SEQUENCE FROM N.A.
RX MEDLINE=97207269; PubMed=9054404;
RA Johnson K.R., Weirick W.C., Zoll W.L., Zhu Y.;
RT "Identification of cDNA clones for the large subunit of eukaryotic
RT translation initiation factor 3. Comparison of homologues from human,
RT Nicotiana tabacum, Caenorhabditis elegans, and Saccharomyces
RT cerevisiae.";
RL J. Biol. Chem. 272:7106-7113(1997).
RL -I- FUNCTION: BINDS TO THE 40S RIBOSOME AND PROMOTES THE BINDING OF
CC METHIONYL-TRNAI AND MRNA.
CC -I- SUBUNIT: EIF-3 IS COMPOSED OF AT LEAST 12 DIFFERENT SUBUNITS.
CC -I- SUBCELLULAR LOCATION: Cytoplasmic.
CC -I- DOMAIN: CONTAINS 1 PCI DOMAIN.
CC -I- PTM: PHOSPHORYLATED.
CC -I- SIMILARITY: BELONGS TO THE EIF3S10 FAMILY.
CC
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; D50929; BA009488.1; -
CC EMBL; U58046; AAB41584.1; -
CC EMBL; U78311; AAB80695.1; -
CC Genew; HGNC:3271; EIF3S10.
CC GK; Q14152; -
CC KIM; 602039; -
CC DR GO; GO:0005852; Eukaryotic translation initiation factor 3. . . TAS.
CC DR GO; GO:0006446; Regulation of translational initiation; TAS.
CC DR InterPro; IPR000717; PCI.

```

```

DR Pfam; PF01399; PCI; 1.
DR SMART; SM00088; PINT; 1.
DR KIRK; KIRK000000; Protein biosynthesis; Repeat; Phosphorylation.
FT Initiation factor; 25 X 10 AA. TANDEN REPEAT OF D-[DE]-D-R-
FT DOMAIN 925 1172 [GP]-[PS]-[RW]-R-[GN]-[AM].
SQ SEQUENCE 1382 AA; 166568 MM; 48501828067BBA CRC64;

Query Match 6.0%; Score 108; DB 1; Length 1382;
Best Local Similarity 21.5%; Pred. No. 2.5;
Matches 68; Conservative 33; Mismatches 118; Indels 98; Gaps 12;

QY 39 QQYSHRGRKQQTTEOREWLSYCGVYVWNSQYPTVRLAFASFDDEDFKMLNGR 98
DB 785 EHRHRLBERKQKEERR-----ITYRKEEERQRRAEQMLERRERRARRA 835
QY 99 PRSGTRAEFEGRVAKKSPFEKGFQRAEVAQVNRRLAEVADSAVDNKKELANGN 158
DB 836 KKEEELR-EYQERVKLEVERKRRQRELEERRRR-----EEERRLGD 880
QY 159 DALRNDAR-----SPFSALRTPSPKE-RNGNHDPSSRKAVYSKHPWS 204
DB 881 SLSKRDSRMDRDSBGTMKQPEADSEMRNGPPEKMRGGERD----- 925
QY 205 GQDRSSSADK---RKYDDP---AFRP---APGTGVMSRDNIPSPSPSGEGRVN 253
DB 926 -EDRSRRRDEPRRLGDDEDRPPLRPDDRVPRGDDDRGPRGPEDEFSRRGADD 984
QY 254 FDGFGAGQTE-----ADADKVTWTHGNHYHAPNGSIGAMHYESKPERNMSGSDPDR 307
DB 985 DRPSWNTDDPPRRRLADDRGNWNRHADDRPPRGL-----DSDR 1026
QY 308 GAVYTFPKSWNTAPD 324
DB 1027 G-----SWRTADE 1034

RESULT 3
ANYB_BACCI STANDARD; PRT; 575 AA.
ID ANYB_BACCI
AC P06547;
DT 01-JAN-1988 (Rel. 06, Created)
DT 01-JAN-1988 (Rel. 06, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Beta-amylase precursor (EC 3.2.1.2) (1,4-alpha-D-glucan
DE maltohydrolase).
DE Bacillus circulans.
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX NCBI_TaxID=1397;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=NCIB 11033;
RX MEDLINE=88260890; PubMed=2455212;
RA Siggens K.W.;
RT "Molecular cloning and characterization of the beta-amylase gene from
RT Bacillus circulans.";
RL Mol. Microbiol. 1:86-91(1987).
RL -I- CATALYTIC ACTIVITY: Hydrolysis of 1,4-alpha-glucosidic linkages in
CC polysaccharides so as to remove successive maltose units from the
CC non-reducing ends of the chains.
CC -I- SUBUNIT: Monomer.
CC -I- SIMILARITY: BELONGS TO FAMILY 14 OF GLYCOSYL HYDROLASES
CC (BETA-AMYLASES).
CC
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; Y00523; CAA68578.1; -
CC PIR; S03745; S03745.

```

CC	use by non-profit institutions as long as its content is in no way
CC	modified and this statement is not removed. Usage by and for commercial
CC	entities requires a license agreement (See http://www.ishb-sib.ch/announce/
CC	or send an email to licenses@sib-sib.ch).
CC	-----
DR	EMBL; M97180; AAB01370.1; -.
DR	PIR; B37271; B37271.
DR	InterPro; IPR001356; Homeobox.
DR	Pfam; PF04611; Aaliphay_MDB; 1.
DR	Pfam; PF00046; homeobox; 1.
DR	ProDom; PD00010; Homeobox; 1.
DR	SMART; SM00389; HOX; 1.
DR	PROSITE; PS00027; HOMEBOX_1; 1.
DR	PROSITE; PS50071; HOMEBOX_2; 1.
DR	Homeobox; DNA-binding; Transcription regulation; Nuclear protein.
Kw	Homeobox; DNA-binding; Transcription regulation; Nuclear protein.
PT	DNA BIND 147 206
PT	DOMAIN 620 660 ARG/LYS-RICH (BASIC).
FT	DOMAIN 677 724 SER-RICH.
SQ	SEQUENCE 926 AA; 10183 MW; 186631422D3E2B CRC64;
Query Match	5.5%; Score 100.5; DB 1; Length 926;
Best Local Similarity	20.5%; Pred. No.5.6;
Matches	71; Conservative 36; Mismatches 122; Indels 117; Gaps 15;
QY	9 PAEPLDMPDPFPRSPSYGAEFTVANNVYRKQCYVSHDGRKQMTBEGREWLTVGCVGT 68
Db	601 PAEIVYR-PDDAPFPVALAEKRAARRRRKKKQAEKRRK----- 640
QY	69 WWSGQYPTNRLAFASFDEDFPKNELKNGRSPSGETRAPEFGYAKESFDEKGFQRAE 128
Db	641 -----DEKRRAREAKQAKKREKQRAGLPRR-SPTLDS-----SRASS 678
QY	129 VASVYMRALFNHSDASVAILDLKKELANGNDALRNEDARSPFYALNTSPFKRRGNH 188
Db	679 VTSASATSKSKRTKRKRSDSASVSS-----AKTP-----SLSTGS-RRSGIS 724
QY	169 DPS--RKK--AVIYSKFPWSGQDRSSAD---KRYGDDPA----- 222
Db	725 MPAPPRNSELPTVASDPLVLTGDKVTYTPPELMAQLFGEDASGLDEPMQSGFSDML 784
QY	223 -FRAPGTGLVMSNDRIIPSPSPBEGVNFVNYWFG--AQREADADKTWTAGNHT 279
Db	785 IFSSCNDGALGNTLVADVMPE-----LGLSDTQLSPDMNTSSMDLS 828
QY	280 A-PWGSILGANHYEESKFRNWSBEGYSDRDGAVITPTPKSWNTAPD 324
Db	829 TQPAASPDSSSETSSMWDN-----LLPQCANTAPD 859
RESULT 5	
JPH3 MOUSE	
ID	STANDARD; PRT; 744 AA.
JPH3 MOUSE	
AC	Q9ET77; Q9EQ22;
DT	26-FEB-2003 (Rel. 41, Created)
RC	28-FEB-2003 (Rel. 41, Last sequence update)
DT	15-SEP-2003 (Rel. 42, Last annotation update)
DE	Uncatophilin 3 (Uncatophilin type 3) (CP-3).
GN	JPH3 OR JP3.
OS	Mus musculus (Mouse).
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
CC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX	NCBI_TaxId=10090;
RN	[1]
RP	SEQUENCE FROM N.A., AND SUBCELLULAR LOCATION.
RC	STRAIN=C57BL/6J; and 129; TISSUE=Brain;
RA	MEDLINE=2040329; PubMed=10949023;
RT	Takeshima H., Komazaki S., Nishi M., Iino M., Kangawa K.;
RL	Uncatophilins: a novel family of junctional membrane complex
RU	proteins".
RU	Mol. Cell 6:11-22(2000).
RN	[2]
RP	FUNCTION, AND TISSUE SPECIFICITY.
RU	MEDLINE=21903756; PubMed=11906164;

RA Nishi M., Hashimoto K., Kuriyama K., Komazaki S., Kano M., Shibata S.,
 RA Takeshima H.;
 RT "Motor discoordination in mutant mice lacking junctophilin type 3";
 RL Biochem. Biophys. Res. Commun. 292:318-324(2002).
 CC -1- FUNCTION: Contributes to the stabilization of the junctional
 CC membrane complexes, which are common to excitable cells and
 CC mediate cross-talk between cell surface and intracellular ion
 CC channels. Probably acts by anchoring the plasma membrane and
 CC endoplasmic reticulum (By similarity). May play an active role in
 CC certain neurons involved in motor coordination.
 CC -1- SUBCELLULAR LOCATION: Type IV membrane protein. Localized
 CC predominantly on the plasma membrane. The transmembrane domain is
 CC anchored in endoplasmic reticulum membrane, while the N-terminal
 CC part associates with the plasma membrane.
 CC -1- TISSUE SPECIFICITY: Specifically expressed in brain. Expressed in
 CC certain populations of neurons but not in glial cells. In
 CC cerebellar sections, it is highly expressed in Purkinje cells,
 CC while it is weakly expressed in granular cells.
 CC -1- DOMAIN: The KORN (membrane occupation and recognition nexus)
 CC repeats contribute to the plasma membrane binding, possibly by
 CC interacting with phospholipids (By similarity).
 CC -1- MISCELLANEOUS: JPH3 deficient mice are viable and fertile but have
 CC defects in balance/motor coordination tasks.
 CC -1- SIMILARITY: BELONGS TO THE JUNCTOPHILIN FAMILY.
 CC -1- SIMILARITY: Contains 8 KORN repeats.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC
 CC -----
 DR EMBL; AB024449; BAB12046.1; -
 DR EMBL; AB024450; BAB20320.1; -
 DR MGD; MGI:1891497; Jpn3.
 DR InterPro; IPR003409; KORN.
 DR Pfam; PF02493; KORN; 8.
 DR SMART; SM00698; KORN; 6.
 KW Endoplasmic reticulum; Transmembrane; Repeat.
 FT DOMAIN 1 723 CYTOPLASMIC (POTENTIAL).
 FT TRANSEM 724 744 ANCHOR FOR TYPE IV MEMBRANE PROTEIN
 FT
 FT REPEAT 15 37 (POTENTIAL).
 FT REPEAT 39 60 MORN 1.
 FT REPEAT 61 82 MORN 2.
 FT REPEAT 83 105 MORN 3.
 FT REPEAT 107 129 MORN 4.
 FT REPEAT 130 152 MORN 5.
 FT REPEAT 152 174 MORN 6.
 FT REPEAT 174 196 MORN 7.
 FT REPEAT 196 218 MORN 8.
 FT DOMAIN 4 143 GLY-RICH.
 FT DOMAIN 143 333 ALA-RICH.
 FT DOMAIN 333 416
 FT SEQUENCE 744 AA; 81229 MW; 3D72AED6A6DA914 CRC64;
 Query Match 5.5%; Score 100; DB 1; Length 744;
 Best Local Similarity 22.0%; Pred. No. 4.7;
 Matches 63; Conservative 35; Mismatches 112; Indels 76; Gaps 14;
 QY 23 SYGARAE-----TVNNYIRKWO-----YSHRDQRKQWTEBEERLTS-- 61
 DB 265 SLSEALAEALVIEDIDITTEITYVGMKDKRSGFGVSGQSDGLKT-----EGEWSNR 319
 QY 62 ---YGCV---GTVWWSGQYPTNRLAFASFEDRFKNELANGRPSEGTREAFEGSVAK 114
 DB 320 RHGYGCMTEPDGTR--EKGKKQWLV-----SGKKNLIPLRASX 358
 QY 115 ESFDEKGFQRRARFASVNMALFNHDSAYIDNLKEIANGDARMDASPRYSAL 174
 DB 359 TREKVDRAVEALERATIAKQKVEIASRISH-SAPKAAALTPAOKQGEAFIARITAK 417
 QY 175 RNTPSFKRNGNHPDRMKAVIYS--KHPWGGQDRS--SSADRKRTGDDPARPAPG 230

DB 418 ESFSEFORENG-----LEYQRPKHQSCDIEVLTSTGPLQGSPELYR--KQTT 466
 QY 231 LVDMSRD-----RNPSPSPS--PGESEFNPFYMGFGATADADKT 270
 DB 467 PSDLTPDSPLOSFPASPSTSTPPAPASRTMAHFSQVSDERS 512
 RESULT 6
 RS22 DROME
 ID RS22 DROME STANDARD; PRT, 943 AA.
 AC Q9V9K7;
 DT 28-FEB-2003 (rel. 41, Created)
 DT 28-FEB-2003 (rel. 41, Last sequence update)
 DT 28-FEB-2003 (rel. 41, Last annotation update)
 DE Aresinle-resistance protein 2 homolog.
 GN C97843.
 OS Drosophila melanogaster (fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Insecta; Pterygota;
 OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC Ephydroidea; Drosophilidae; Drosophila.
 OX NCBI_TaxId=7227;
 [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Berkeley;
 RX MEDLINE=20196006; PubMed=10731132;
 RA Adams M.D., Celinker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
 RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galie R.F.,
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
 RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
 RA Brandon R.C., Rogers Y.-H.C., Blazey R.G., Champe M., Pfeiffer B.D.,
 RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
 RA Abril J.F., Abmayyan A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
 RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
 RA Beeson K.V., Betts P.V., Berman B.P., Bhandari D., Bolshakov S.,
 RA Bokora D., Botchan M.R., Bouck J., Brockstein P., Brottier P.,
 RA Butts K.C., Busan D.A., Butler H., Cadieu E., Center A., Chandra I.,
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
 RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
 RA Dodson K., Dou P.L.E., Downes M., Dugan-Bocsa S., Dunkov B.C., Dunn P.,
 RA Duhail K.J., Evangelista C.C., Ferraz C., Ferlita S., Fleischmann W.,
 RA Folsler C., Gabrielian A.E., Gang N.S., Gelbart W.M., Glasser K.,
 RA Glodek A., Gong P., Gottlieb J.H., Gu Z., Guan P., Harris M.,
 RA Harris N.L., Harvey D., Heiman T.D., Hernandez J.R., Houck J.,
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Idegawa C.,
 RA Jaisl M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
 RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
 RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Mostrelet A.,
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
 RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
 RA Reinert K., Remington K., Saunders R.D.C., Schaefer F., Shen H.,
 RA Shue B.C., Siden-Kimms I., Simpson M., Skupski M.P., Smith T.,
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
 RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
 RA Wang Z.-Y., Wasserman D.A., Weinstock G.M., Weissbach J.,
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
 RA Ye J., Yen R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
 RA Gibbs R.A., Myers E.W., Rubin G.W., Venter J.C.;
 RT "The genome sequence of Drosophila melanogaster";
 RL Science 287:2185-2195(2000).
 [2]
 REVISIONS, AND ALTERNATIVE SPLICING.
 RP STRAIN=Berkeley;
 RX MEDLINE=22426069; PubMed=12537572;
 RA Maier S., Crosby W.A., Mungall C.J., Matthews B.B., Campbell K.S.,
 RA Hradecky P., Huang Y., Kaminker J.S., Milburn G.H., Prochownik S.E.,
 RA Smith C.D., Tupy J.L., Whitfield E.J., Bayraktaroglu L., Berman B.P.,
 RA Betencourt B.R., Celinker S.E., de Grey A.D.N.J., Drysdale R.A.,
 RA Harris N.L., Richter J., Russo S., Schroeder A.J., Shu S.Q.,

RA Stapleton M., Yamada C., Ashburner M., Gelbart W.M., Rubin G.M.,
 RA Lewis S.B.;
 RT "Annotation of the Drosophila melanogaster euchromatic genome: a
 RT systematic review.";
 RL Genome Biol. 3:RESEARCH0083.1-RESEARCH0083.22(2002).
 CC -!- ALTERNATIVE PRODUCTS;
 CC Event=Alternative splicing; Named isoforms=2;
 CC Name=long;
 CC IsoId=O9V9K7-1; Sequence=Displayed;
 CC Name=short;
 CC IsoId=O9V9K7-2; Sequence=VSP_000327;
 CC -!- SIMILARITY: BELONGS TO THE ARS2 FAMILY.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

DR EMBL, AE003784; AAM68343.1;
 DR EMBL, AE003784; AAM68345.1;
 DR Flybase, FB99003062; CG7843.
 DR Pfam, PF04959; ARS2; 1.
 KM Hypothetical protein; Alternative splicing.
 FT VARSPLIC 47 Missing (in isoform short).
 FT /FTD=VSP_000327.
 FT SEQUENCE 943 AA; 107221 MW; 0C1AF03E02E8AB08 CRC64;

Query Match 5.5%; Score 99.5; DB 1; Length 943;
 Best Local Similarity 18.3%; Pred. No. 6.9;
 Matches 69; Conservative 43; Mismatches 144; Indels 121; Gaps 13;

QY 10 AEPLDRPDYRSGYGAATVYNNYIKQVYSHDCKQOMTEBQREMLSGCVGW 69
 DB 311 ADPSTGRKVRP-----VNSGEMVDDDAENSAKKELABDSD----- 351
 QY 70 VNSQYPTNELAFASFEDEFRKNEIKNGRPR-----SGETRAEFGRVAKESFDEE 120
 DB 352 --SDSK-----EDKQINKKTKYKGRKNSDDSSSESSSSSDEKIKEXYDVE 399
 QY 121 KGPRAEIVASVWNRALENADESAVYDNKKEIANGNDALNDDASPRYSALRTPEF 180
 DB 400 DGLFAEQKTEAEKDRBEATKAKQGPQPKLDEDEGENETEPKGLDSKINTYEBIDWTKS 459
 QY 181 KE-----RNGGNHDPSEK-----KAVIYSKHFWSGDQ-----SSADKXKGDPAFR 224
 DB 460 PEISNPRIKXTDNGDSKTEBEDGKRPV-----CKDKVYETETIDLDKVKQGPRAH 512
 QY 225 PAPTGTIVNDRDRIIPERSPT-SPGEGFVN-----FDYGFPAQTE 264
 DB 513 RTSSIFL-----RNLAPEITSEIEVACNRFSGYLVAIADPVERRWRRGWITFMD 566
 QY 265 ADADKVTWTHGN-----HYAPNGS.GANHYVES----- 293
 DB 567 VNIKELTQGNQRLDCEMAIYVRPLSRVAPANGITAHQVYRSDIKCAKIALND 626
 QY 294 -KFNMSSEGYSDPDRGA 309
 DB 627 EKFLMAEGPKDSDNSGA 643

RESULT 7
 CCAA_RABIT STANDARD; PRT; 2424 AA.
 ID CCAA_RABIT
 AC P27884; P27883;
 RT 01-JUN-1993 (Rel. 26, Created)
 RT 01-JUN-1993 (Rel. 26, Last sequence update)
 RT 15-SEP-2003 (Rel. 42, Last annotation update)
 RT Voltage-dependent P/Q-type calcium channel alpha-1A subunit (Calcium
 RT channel, L type, alpha-1 polypeptide isoform 4) (Brain calcium channel
 RT 1) (B1).

GN CACNA1A OR CACNA1A4 OR CACNA4 OR CACNA3.
 OS Oryctolagus cuniculus (Rabbit).
 CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 CC Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.
 CX NCBI_Taxid=9986;
 RN [1]
 RP SEQUENCE FROM N.A., AND ALTERNATIVE SPLICING.
 RC TISSUE=Brain;
 RX MEDLINE=9118710; Pubmed=184923;
 RA Mori Y., Friedrich T., Kim M.-S., Mikami A., Nakai J., Ruth P.,
 RA Bosse E., Hoffmann F., Flockeitz V., Furcht T., Mikoshiba K.,
 RA Imoto K., Tanabe T., Numa S.;
 RT "Primary structure and functional expression from complementary DNA
 RT of a brain calcium channel.";
 RL Nature 350:398-402(1991).
 RN [2]
 RP BETA-SUBUNIT BINDING DOVAIN AND MUTAGENESIS.
 RX MEDLINE=94150724; Pubmed=7509046;
 RA Pragnell M., de Waard M., Mori Y., Tanabe T., Snutch T.P.,
 RA Campbell K.P.;
 RT "Calcium channel beta-subunit binds to a conserved motif in the I-II
 RT cytoplasmic linker of the alpha 1-subunit.";
 RL Nature 368:67-70(1994).
 CC -!- FUNCTION: VOLTAGE-SENSITIVE CALCIUM CHANNELS (VSCC) MEDATE THE
 CC ENTRY OF CALCIUM IONS INTO EXCITABLE CELLS AND ARE ALSO INVOLVED
 CC IN A VARIETY OF CALCIUM-DEPENDENT PROCESSES, INCLUDING MUSCLE
 CC CONTRACTION, HORMONE OR NEUROTRANSMITTER RELEASE, GENE EXPRESSION,
 CC CELL MOTILITY, CELL DIVISION AND CELL DEATH. THE ISOFORM ALPHA-1A
 CC GIVES RISE TO P AND/OR Q-TYPE CALCIUM CURRENTS (P/Q-TYPE CALCIUM
 CC CHANNELS BELONG TO THE "HIGH-VOLTAGE ACTIVATED" (HVA) GROUP AND
 CC ARE BLOCKED BY THE FUNNEL TOXIN (FTX) AND BY THE OMEGA-AGATOXIN-
 CC IVA (OMEGA-AEA-IVA). THEY ARE HOWEVER INSENSITIVE TO
 CC DIHYDROPYRIDINES (DHP), AND OMEGA-CONOTOXIN-GVIA (OMEGA-CTX-GVIA).
 CC -!- SUBUNIT: VOLTAGE-DEPENDENT CALCIUM CHANNELS ARE MULTISUBUNIT
 CC COMPLEXES, CONSISTING OF ALPHA-1, ALPHA-2, BETA AND DELTA SUBUNITS
 CC IN A 1:1:1:1 RATIO. THE CHANNEL ACTIVITY IS DIRECTED BY THE PORE-
 CC FORMING AND VOLTAGE-SENSITIVE GENERATE VOLTAGE-SENSITIVE CALCIUM
 CC SUBUNIT IS SUFFICIENT TO GENERATE VOLTAGE-SENSITIVE CALCIUM
 CC CHANNEL ACTIVITY. THE AUXILIARY SUBUNITS BETA AND ALPHA-2/DELTA
 CC LINKED BY A DISULFIDE BRIDGE REGULATE THE CHANNEL ACTIVITY.
 CC -!- SUBCELLULAR LOCATION: Integral membrane protein.
 CC -!- ALTERNATIVE PRODUCTS:
 CC Event=Alternative splicing; Named isoforms=5;
 CC Name=BI-2; Synonyms=1A-2;
 CC IsoId=P27884-1; Sequence=Displayed;
 CC Name=BI-1; Synonyms=1A-1;
 CC IsoId=P27884-2; Sequence=VSP_000879, VSP_000880;
 CC Name=CBP101; Synonyms=CBP109;
 CC IsoId=P27884-3; Sequence=VSP_000878;
 CC Name=CBP103;
 CC IsoId=P27884-4; Sequence=VSP_000877;
 CC Name=CBP107;
 CC IsoId=P27884-5; Sequence=VSP_000876;
 CC -!- TISSUE SPECIFICITY: BRAIN-SPECIFIC. PURKINJE CELLS CONTAIN
 CC PREDOMINANTLY P-TYPE VSCC, THE Q-TYPE BEING A PROMINENT CALCIUM
 CC CURRENT IN CEREBELLAR GRANULE CELLS.
 CC -!- DOMAIN: EACH OF THE FOUR INTERNAL REPEATS CONTAINS FIVE
 CC HYDROPHOBIC TRANSMEMBRANE SEGMENTS (S1, S2, S3, S5, S6) AND ONE
 CC POSITIVELY CHARGED TRANSMEMBRANE SEGMENT (S4). S4 SEGMENTS
 CC PROBABLY REPRESENT THE VOLTAGE-SENSOR AND ARE CHARACTERIZED BY A
 CC SERIES OF POSITIVELY CHARGED AMINO ACIDS AT EVERY THIRD POSITION.
 CC -!- SIMILARITY: BELONGS TO THE CALCIUM CHANNEL ALPHA-1 SUBUNITS
 CC FAMILY.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

DR EMBL, X57477; CAA40715.1; -.

[illegible]

FT	DOMAIN	1012	1017	POLY-ARG.
FT	DOMAIN	2219	2227	POLY-HIS.
FT	DOMAIN	2242	2246	POLY-ARG.
FT	DOMAIN	2288	2297	POLY-ARG.
FT	DOMAIN	2298	2301	POLY-GLY.
FT	DOMAIN	2312	2377	POLY-PRO.
FT	DOMAIN	2411	2416	POLY-GLY.
FT	DOMAIN	383	400	BINDING TO THE BETA SUBUNIT.
FT	SITE	318	318	CALCIUM ION SELECTIVITY AND PERMEABILITY
FT	SITE	668	668	(BY SIMILARITY).
FT	SITE	1469	1469	CALCIUM ION SELECTIVITY AND PERMEABILITY
FT	SITE	1765	1765	(BY SIMILARITY).
FT	SITE	1765	1765	CALCIUM ION SELECTIVITY AND PERMEABILITY
FT	SITE	1765	1765	(BY SIMILARITY).
FT	MOD_RES	1831	1831	PHOSPHORYLATION (BY PKA) (POTENTIAL).
FT	CA_BIND	1849	1860	BY SIMILARITY.
FT	CARBOHYD	283	283	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	1665	1665	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	VARSPPLIC	772	1051	Missing (in isoform CBP107).
FT	VARSPPLIC	772	1120	/ftid=VSP_000876.
FT	VARSPPLIC	772	1120	Missing (in isoform CBP103).
FT	VARSPPLIC	1857	1884	/ftid=VSP_000877.
FT	VARSPPLIC	1857	1884	LYRPMYALHMPPLPGGLKXNCPARVAV -> HYKDMYSL
FT	VARSPPLIC	1857	1884	RVISPLPLGLKXKPPHVCAG (in isoform CBP101).
FT	VARSPPLIC	2230	2273	/ftid=VSP_000878.
FT	VARSPPLIC	2230	2273	RGPRGVSPGTSARRRRRGVAVVAPARAPALARRAPAP
FT	VARSPPLIC	2230	2273	AKL -> PAADKERYGPQDRPDHGGARARDDQWRSPS
FT	VARSPPLIC	2230	2273	EGRRHTTHRQ (in isoform BI-1).
FT	VARSPPLIC	2274	2424	/ftid=VSP_000879.
FT	VARSPPLIC	2274	2424	Missing (in isoform BI-1).
FT	VARSPPLIC	2274	2424	/ftid=VSP_000880.
FT	VARSPPLIC	419	419	MISSING (IN ISOFORM CBP315).
FT	VARSPPLIC	877	877	A -> T (IN ISOFORM CBS).
FT	VARSPPLIC	1104	1104	S -> N (IN ISOFORM CBS).
FT	VARSPPLIC	386	386	E->S. REDUCED BETA-SUBUNIT INTERACTION.
FT	VARSPPLIC	389	389	L->H. REDUCED BETA-SUBUNIT INTERACTION.
FT	VARSPPLIC	392	392	Y->S. REDUCED BETA-SUBUNIT INTERACTION.
FT	VARSPPLIC	400	400	E->A. NO EFFECT ON BETA-SUBUNIT
FT	VARSPPLIC	400	400	INTERACTION.
FT	VARSPPLIC	400	400	INTERACTION.
FT	VARSPPLIC	400	400	INTERACTION.
FT	VARSPPLIC	400	400	INTERACTION.
FT	VARSPPLIC	400	400	INTERACTION.
FT	VARSPPLIC	400	400	INTERACTION.
FT	VARSPPLIC	400	400	INTERACTION.
FT	VARSPPLIC	400	400	INTERACTION.
FT	VARSPPLIC	400	400	INTERACTION.
FT	VARSPPLIC	400	400	INTERACTION.
FT	VARSPPLIC	400	400	INTERACTION.
FT	VARSPPLIC	400	400	INTERACTION.
FT	VARSPPLIC	400	400	INTERACTION.
FT	VARSPPLIC	400	400	INTERACTION.
FT	VARSPPLIC	400	400	INTERACTION.
FT	VARSPPLIC	400	400	INTERACTION.
FT	VARSPPLIC	400	400	INTERACTION.
FT	VARSPPLIC	400	400	INTERACTION.
FT	VARSPPLIC	400	400	INTERACTION.
FT	VARSPPLIC	400	400	INTERACTION.
FT	VARSPPLIC	400	400	INTERACTION.
FT	VARSPPLIC	400	400	INTERACTION.
FT	VARSPPLIC	400	400	INTERACTION.
FT	VARSPPLIC	400	400	INTERACTION.
FT	VARSPPLIC	400	400	INTERACTION.
FT	VARSPPLIC	400	400	INTERACTION.
FT	VARSPPLIC	400	400	INTERACTION.
FT	VARSPPLIC	400	400	INTERACTION.
FT	VARSPPLIC	400	400	INTERACTION.
FT	VARSPPLIC	400	400	INTERACTION.
FT	VARSPPLIC	400	400	INTERACTION.
FT	VARSPPLIC	400	400	INTERACTION.
FT	VARSPPLIC	400	400	INTERACTION.
FT	VARSPPLIC	400	400	INTERACTION.
FT	VARSPPLIC	400	400	INTERACTION.
FT	VARSPPLIC	400	400	INTERACTION.
FT	VARSPPLIC	400	400	INTERACTION.
FT	VARSPPLIC	400	400	INTERACTION.
FT	VARSPPLIC	400	400	INTERACTION.
FT	VARSPPLIC	400	400	INTERACTION.
FT	VARSPPLIC	400	400	INTERACTION.
FT	VARSPPLIC	400	400	INTERACTION.
FT	VARSPPLIC	400	400	INTERACTION.
FT	VARSPPLIC	400	400	INTERACTION.
FT	VARSPPLIC	400	400	INTERACTION.
FT	VARSPPLIC	400	400	INTERACTION.
FT	VARSPPLIC	400	400	INTERACTION.
FT	VARSPPLIC	400	400	INTERACTION.
FT	VARSPPLIC	400	400	INTERACTION.
FT	VARSPPLIC	400	400	INTERACTION.
FT	VARSPPLIC	400	400	INTERACTION.
FT	VARSPPLIC	400	400	INTERACTION.
FT	VARSPPLIC	400	400	INTERACTION.
FT	VARSPPLIC	400	400	INTERACTION.
FT	VARSPPLIC	400	400	INTERACTION.
FT	VARSPPLIC	400	400	INTERACTION.
FT	VARSPPLIC	400	400	INTERACTION.
FT	VARSPPLIC	400	400	INTERACTION.
FT	VARSPPLIC	400	400	INTERACTION.
FT	VARSPPLIC	400	400	INTERACTION.
FT	VARSPPLIC	400	400	INTERACTION.
FT	VARSPPLIC	400	400	INTERACTION.
FT	VARSPPLIC	400	400	INTERACTION.
FT	VARSPPLIC	400	400	INTERACTION.
FT	VARSPPLIC	400	400	INTERACTION.
FT	VARSPPLIC	400	400	INTERACTION.
FT	VARSPPLIC	400	400	INTERACTION.
FT	VARSPPLIC	400	400	INTERACTION.
FT	VARSPPLIC	400	400	INTERACTION.
FT	VARSPPLIC	400	400	INTERACTION.
FT	VARSPPLIC	400	400	INTERACTION.
FT	VARSPPLIC	400	400	INTERACTION.
FT	VARSPPLIC	400	400	INTERACTION.
FT	VARSPPLIC	400	400	INTERACTION.
FT	VARSPPLIC	400	400	INTERACTION.
FT	VARSPPLIC	400	400	INTERACTION.
FT	VARSPPLIC	400	400	INTERACTION.
FT	VARSPPLIC	400	400	INTERACTION.
FT	VARSPPLIC	400	400	INTERACTION.
FT	VARSPPLIC	400	400	INTERACTION.
FT	VARSPPLIC	400	400	INTERACTION.
FT	VARSPPLIC	400	400	INTERACTION.
FT	VARSPPLIC	400	400	INTERACTION.
FT	VARSPPLIC	400	400	INTERACTION.
FT	VARSPPLIC	400	400	INTERACTION.
FT	VARSPPLIC	400	400	INTERACTION.
FT	VARSPPLIC	400	400	INTERACTION.
FT	VARSPPLIC	400	400	INTERACTION.
FT	VARSPPLIC	400	400	INTERACTION.
FT	VARSPPLIC	400	400	INTERACTION.
FT	VARSPPLIC	400	400	INTERACTION.
FT	VARSPPLIC	400	400	INTERACTION.
FT	VARSPPLIC	400	400	INTERACTION.
FT	VARSPPLIC	400	400	INTERACTION.
FT	VARSPPLIC	400	400	INTERACTION.
FT	VARSPPLIC	400	400	INTERACTION.
FT	VARSPPLIC	400	400	INTERACTION.
FT	VARSPPLIC	400	400	INTERACTION.
FT	VARSPPLIC	400	400	INTERACTION.
FT	VARSPPLIC	400	400	INTERACTION.
FT	VARSPPLIC	400	400	INTERACTION.
FT	VARSPPLIC	400	400	INTERACTION.
FT	VARSPPLIC	400	400	INTERACTION.
FT	VARSPPLIC	400	400	INTERACTION.
FT	VARSPPLIC	400	400	INTERACTION.
FT	VARSPPLIC	400	400	INTERACTION.
FT	VARSPPLIC	400	400	INTERACTION.
FT	VARSPPLIC	400	400	INTERACTION.
FT				

```

CX NCBI_TaxID=10090;
RN [1]
RN SEQUENCE FROM N.A.
RA MEDLINE=9313824; PubMed=8421688;
RA Anderson S.K., Gallinger S., Roder J., Frey J., Young H.A.,
RA Ortolano J.R.;
RA "A cyclophilin-related protein involved in the function of natural
RT killer cells.";
RL Proc. Natl. Acad. Sci. U.S.A. 90:542-546(1993).
RN [2]
RP REVISIONS TO C-TERMINUS.
RC STRAIN=BALB/c; TISSUE=Blood;
RA Anderson S.K.;
RL Submitted (OCT-1999) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: COMPONENT OF A PUTATIVE TUMOR-RECOGNITION COMPLEX.
CC INVOLVED IN THE FUNCTION OF NK CELLS.
CC -1- SIMILARITY: Contains 1 cyclophilin-like p1ase domain.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; L04289; AAA37500.2; ALT_INIT.
DR HSSP; Q27450; 1A33.
DR MGD; MGI:97346; NKT.
DR InterPro; IPR002130; CSA_P1ase.
DR Pfam; PF00160; pro_isomerase; 1.
DR PROSITE; PS00170; CSA_P1ase_1; 1.
DR PROSITE; PS50072; CSA_P1ase_2; 1.
KM Cyclosporin; Isomerase; Rotamase; Repeat; Transmembrane.
FT DOMAIN 1 176 P1ase; CYCLOPHILIN-TYPE.
FT DOMAIN 222 241 ARG/LYS-RICH (BASIC).
FT DOMAIN 422 459 ARG/LYS-RICH (BASIC).
FT DOMAIN 964 1003 ARG/LYS-RICH (BASIC).
FT DOMAIN 198 273 ARG/SER-RICH.
FT DOMAIN 468 565 ARG/SER-RICH.
FT DOMAIN 658 812 ARG/SER-RICH.
FT DOMAIN 1303 1453 ARG-SER TANDWM REPEAT-RICH.
SQ SEQUENCE 1453 AA; 163439 MW; DFL173F814B283E CRC64;
Query March 5.4%; Score 98.5; DB 1; Length 1453;
Best Local Similarity 17.7%; Pred. No. 14; Indels 195; Gaps 21;
Matches 88; Conservative 63; Mismatches 152;
QY 1 DSDDRVTPP--AEPDRKVPDPYRPSYRA-----ETVNNYTRKQQ-- 40
DB 355 DDSSSETPPHKKEBQRL-RAYRPSGKWKSGDKLSDPCSRMDEKSLQSRSMVYNG 413
QY 41 VYS-----HRDG-----RKQMTBEDEWLSGCVGVTVN 71
DB 414 YSLSLSTARSDDGHNKRRKKEKFKKAKKQKQGRHRRQTKRR-----IWPDELE 466
QY 72 SGQVPTNRL-----APASPEDRFNNEKKNRPSGRTFAEPGRVAKH 115
DB 467 PSRPFTRKMSCTVRRRSRRASSSSHSSKRWKSDDDDSASTHSRSTY--RSKSH 524
QY 116 SFDEKGFQFARREVAQVNRALENAHDESA-----YLDNWKELANGDALRNDAR 167
DB 525 SRSSRSRSSSRRAVSKSSRSRLNNSKSSSSRSRSPRTSISPKXQAQSEKPKVTEPPR 584
QY 168 -----SPFYALP----- 175
DB 585 PSYQNGNVLVQPAENIPVPLSDSPSPRPMKPGQKWKMSYRIQEMZAKTTHLPV 644
QY 176 -----NTSEFKERNGNHDPSRKAKVITSKHPWGGQSRSSADKRYGDPDA 222
DB 645 QSTYSLTNITATVSSSSYHKR---EKPSBDSGASIKY---SPRSSGSSGR--SSSKS 694
QY 223 FRAPATGLVWMSRDNIP-----RSPT-----SPGEGFVNFQYGFQAQTEADAD---- 268

```

```

DB 695 SRSRSSRSYTRSRSRSLPTSRSLSPSSRSHSPNKYSDGQSHSRSSSYTVSSDDGR 754
QY 269 -----KTVTHGNHYHAPNAGSLAMETV-----YESKFRWSGYSDFD-----RGAVV 311
DB 755 AMFRNKKKSVTSHKRRNSERTLHAKIVRGREKSKHRRKYSRSLDYSDDQSHV 814
QY 312 ITFIPKSNWTAPDKVKQG 329
DB 815 QVY-----SAPEKEXQG 826
RESULT 9
PHOE_SALTY
ID PHOE_SALTY STANDARD; PRT; 350 AA.
AC P30705;
DT 01-APR-1993 (Rel. 25, Created)
DT 01-APR-1993 (Rel. 25, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Outer membrane pore protein E precursor.
GN PHOE OR STM0320.
OS Salmonella typhimurium.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Salmonella.
OX NCBI_TaxID=602;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=LT2;
RX MEDLINE=93083994; PubMed=1280609;
RA Splierings G., Elders R., van Litch B., Hofstra H., Tommassen J.;
RT "Characterization of the Salmonella typhimurium phoB gene and
RT development of Salmonella-specific DNA probes.";
RL Gene 122:45-52(1992).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=LT2 / SCSG412 / ATCC 700720;
RX MEDLINE=21534948; PubMed=11677609;
RA McClelland M., Sanderson K.E., Spliegh J., Clifton S.W., Latreille P.,
RA Courtney L., Portwell L.S., Ali J., Dante M., Du F., Hou S., Layman D.,
RA Leonard S., Nguyen C., Scott K., Holmes A., Grewal N., Mulvaney E.,
RA Ryan E., Sun H., Flores L., Miller W., Stoneking T., Nhan M.,
RA Waterston R., Wilson R.K.;
RT "Complete genome sequence of Salmonella enterica serovar Typhimurium
RT LT2.";
RL Nature 413:852-856(2001).
CC -1- FUNCTION: THIS IS ONE OF THE PROTEINS INDUCED WHEN CELLS ARE
CC GROWN UNDER PHOSPHATE LIMITATION. ITS PROTEIN POSE IS
CC PARTICULARLY EFFICIENT IN THE UPTAKE OF INORGANIC PHOSPHATE,
CC PHOSPHORYLATED COMPOUNDS, AND SOME OTHER NEGATIVELY CHARGED
CC SOLUTES.
CC -1- SUBUNIT: Homotrimer.
CC -1- SUBCELLULAR LOCATION: Integral membrane protein. Outer membrane.
CC -1- SIMILARITY: BELONGS TO THE OMPC/PHOE FAMILY OF PORINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X68023; CA448164.1; -.
DR EMBL; AE008709; AAL19276.1; -.
DR PIR; S25525; S25525.
DR HSSP; P02932; 1PHO.
DR StyGene; SGI0291; phoB.
DR InterPro; IPR003229; OMP_2.
DR InterPro; IPR001702; Porin Gram-ve.
DR Pfam; PF00267; Gram-ve porins; 1.
DR PRINTS; PR00182; ECOLNEIPORIN.
DR ProDom; PD000808; OMP_2; 1.
DR PROSITE; PS00576; GRAM_NEG_PORIN; 1.

```

KW Outer membrane; Transmembrane; Porin; Signal; Complete proteome.

FT STGNL 1 20 OUTER MEMBRANE PORE PROTEIN E.
SQ CHAIN 21 350
SQ SEQUENCE 350 AA; 38762 MW; 9DLB3355AF59877 CRC64;

Query Match 5.4%; Score 98; DB 1; Length 350;

Best Local Similarity 20.8%; Pred. No. 2.6;

Matches 64; Conservative 35; Mismatches 116; Indels 92; Gaps 13;

QY YGRATVNNVYIRKQGVYSHDGRKQMTTEOREWLSYGVGVWVNS---GQPTN 78

DB 34 YGKVKAM-----HYWSDYSDKDG-----DOSYRFFKFEKTEIINDQLTGKRWAE 79

QY RLAFASFDDEDFKRLKNGSPSGETRAFFGRVAKD--SFDEKSGFQRAEYASVWRA 136

DB 80 ---FASNKAEESDSQK-----TRAFMGLKLTGIGTFDGRGLGALIDVAKWDMF 128

QY 137 LBNADBSAYLDNKKELANGDALRNEDAPSPYSALNTSPFKERNQGNHDSRMKAV 196

DB 129 PEPGDSQAQTDNMTWRASGLATVRND---PFGIVDGLDITLQYQKNEEDRD--- 179

QY 197 TSKKFFNSGQDRSSAPKRYGPDPAFPAAGTGLVMSNDKRIKPSFGSGRTVNDY 256

DB 180 -VKQKNGDGFGTSVSYD--FGSSD-FAVSAAYTLSPRTSEQNLQRRGT----- 223

QY 257 GWFGAQTADADKTVWTHGNHYHAPNGSLGAMHYESKFFNMWEGYSDPFGAYITFIP 316

DB 224 ---GDKARA-----WATGVKYVA-----NDIYATFYS 248

QY 317 KSMWTA 323

DB 249 ETRNMT 255

RESULT 10

ID Bx42_DROME STANDARD; PRT; 547 AA.

AC P39736; Q9M390; 01-FEB-1995 (Rel. 31, Created)

DT 01-FEB-1995 (Rel. 31, Last sequence update)

DT 15-SEP-2003 (Rel. 42, Last annotation update)

DE Puff specific protein Bx42.

GN Bx42 OR CG8264.

OS Drosophila melanogaster (Fruit fly).

OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;

OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;

OC Ephydroidea; Drosophilidae; Drosophila.

OC NCBI_TaxID=7227;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=Oregon-R; TISSUE=Embryo;

RX MEDLINE=93048378; PubMed=1424966;

RA Wieland C., Mann S., von Besser H., Saunweber H.;

RT "The Drosophila nuclear protein Bx42, which is found in many puffs on

polytene chromosomes, is highly charged.";

RL Chromosoma 101:517-525(1992).

RN [2]

RP SEQUENCE FROM N.A.

RC STRAIN=Berkley;

RX MEDLINE=20196006; PubMed=10731132;

RA Adams M.D., Ceiniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,

RA Amaratilake P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,

RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,

RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.A.,

RA Brandon R.C., Rogers Y.-H.C., Blazey R.G., Champe M., Pfeiffer B.D.,

RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,

RA Abell J.P., Agrevari A., An H.-J., Andrews-Pfannkuch C., Baldwin D.,
RA Balow R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Bense P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Botkova D., Botchan M.R., Bouck J., Brokstein P., Brodter P.,
RA Butts K.C., Busan D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablo B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Foster C., Gabriellian A.E., Gang N.S., Gelbart W.M., Glasser K.,
RA Glodex A., Gong F., Gottell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibsen C.,
RA Jalali M., Kalush F., Karpen G.F., Ke Z., Kamsion J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasio P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Maitai B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Mishina N.V., Mobarry C., Morris J., Koshire D.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Mizny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusslein D.R., Pacleab J.M.,
RA Palazolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Relbert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svitzkas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.,
RA "The genome sequence of Drosophila melanogaster.";
Science 287:2185-2195(2000).
[3]
RP REVISIONS.
RC STRAIN=Berkley;
RX MEDLINE=2242606; PubMed=12537572;
RA Mista S., Crosby W.A., Murgall C.J., Matthews B.B., Campbell K.S.,
RA Hradecky P., Huang Y., Kaminker J.S., Millburn G.H., Prochuk S.E.,
RA Smith C.D., Tupy J.L., Whitfield E.J., Bayraktaroglu L., Berman B.P.,
RA Bettencourt B.R., Ceiniker S.E., de Grey A.D.N.J., Drysdale R.A.,
RA Harris N.L., Richter J., Russo S., Schroeder A.J., Shu S.O.,
RA Stapleton M., Yamada C., Ashburner M., Gelbart W.M., Rubin G.M.,
RA Lewis S.E.,
RT "Annotation of the Drosophila melanogaster euchromatic genome: a
systematic review.";
RL Genome Biol. 3:RESEARCH0083.1-RESEARCH0083.22(2002).
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN=Berkley; TISSUE=Embryo;
RX MEDLINE=2242606; PubMed=12537569;
RA Stapleton M., Carlson J.W., Brokstein P., Yu C., Champe M.,
RA George R.A., Guarni H., Krommiller B., Pacleab J.M., Park S., Wan K.H.,
RA Rubin G.M., Kaminker S.E.,
RT "A Drosophila full-length cDNA resource.";
RL Genome Biol. 3:RESEARCH0080.1-RESEARCH0080.8(2002).
CC - FUNCTION. MAY PLAY A ROLE IN CHROMATIN STRUCTURE AND FUNCTION.
CC - SUBCELLULAR LOCATION: Nuclear.
CC - DEVELOPMENTAL STAGE: TWO TRANSCRIPTS ARE DETECTED OF SIZES 1.9 AND
CC 2.2 KB. BOTH ARE DETECTED SOON AFTER FERTILIZATION AND SHOW
CC RELATIVELY CONSTANT EXPRESSION DURING THE FIRST 2/3 OF
CC EMBRYOGENESIS. IN 0-3 HR EMBRYOS, THE SMALLER TRANSCRIPT IS
CC PREDOMINANT AND THE LEVELS OF THE TWO TRANSCRIPTS ARE SOMEWHAT
CC REDUCED AT THE LATER STAGES OF DEVELOPMENT, BUT THEY ARE FOUND IN
CC APPROXIMATELY CONSTANT AMOUNTS DURING LARVAL, PUPAL AND ADULT
CC STAGES. THE SMALLER TRANSCRIPT IS SUSPECTED TO BE A MATERNAL
CC TRANSCRIPT.
CC - SIMILARITY: BELONGS TO THE SNW FAMILY.
CC This SNW-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: X64536; CAA45834.1; -
CC EMBL: AB003446; AAF46444.2; -
CC EMBL: A111364; AAM29369.1; -
CC PIR: A56575; A56575.

DR FlyBase: FBgn004856; Bx42.
 DR InterPro: IPR004015; SKIP_SNM.
 DR Pfam: PF02731; SKIP_SNM; 1.
 KW Nuclear protein; DNA-binding.
 FT DOMAIN 177 343 SNM.
 FT DOMAIN 383 462 SH2-LIKE DOMAIN.
 FT DOMAIN 209 236 PRO-RICH.
 FT DOMAIN 125 155 ASH/GU-RICH (ACIDIC).
 FT DOMAIN 300 400 ASH/GU-RICH (ACIDIC).
 SQ SEQUENCE 547 AA; 61156 MW; 01399EA231C9D557 CRC64;
 Query Match 5.4%; Score 98; DB 1; Length 547;
 Best Local Similarity 22.6%; Pred. No. 4.5;
 Matches 73; Conservative 38; Mismatches 116; Indels 96; Gaps 19;
 QY 44 HDGKQOQTEOREEMLSYGCVGWVWNSGOYPTNRLA-----FAEFD 86
 DB 234 HSPSKVTV-KEQKEMKIPPCIS-NWKNAKVTTPDLKRLAAGRLQGVHNEKFA 291
 QY 87 E-----DFKNELKNRPSRGSETRAFEGRVAKSPDEKGFQAREVASVNNRLENVAH 141
 DB 292 EALYIADRKARE-----AVEARSQLKELAKK--EKEKEDMLRMA-----QEAR 335
 QY 142 DSSATLDLKKELANGDALRNDARSPFYALNTSPFRNGNHDSPKAVIYSKH 201
 DB 336 EERRA-----GLRNEEAEPGSGATGS-EVEREN----- 363
 QY 202 FMSGQDRSSADKKYCDPAPAP-CTGLVDSRPNTPRS-----FTSPGPFYNF 254
 DB 364 -----DLAEQRQRQCDRLQRADEKRSKIQERERDLSQIALGPAKANGETLP 418
 QY 255 DVGWFGAQTADA--DKTWTHGNHYAP--NGSIGAHVYESKFRNWSGY-SDPDR 307
 DB 419 DQRLNNTTKMGDSYGDDEAY--NYYDKPMDSNITLGA-HYRPSKQSDSNYGDLDA 474
 QY 308 GAVVTFTIP-KSNWTPADRKQK 329
 DB 475 IYNTKRFVFDKQFSGASKAAG 497
 RESULT 11
 NKCR_HUMAN STANDARD; PRT; 1462 AA.
 ID NKCR_HUMAN
 AC P30414;
 DT 01-APR-1993 (Rel. 25, Created)
 DT 30-MAY-2000 (Rel. 39, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE NK-tumor recognition protein (Natural killer cells cyclophilin-related protein) (NK-TR protein).
 GN NKTR.
 OS Homo sapiens (Human).
 OC Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 CX NCBI_Taxid=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Blood;
 RX MEDLINE=93133824; PubMed=8421688;
 RA Anderson S.K., Gallinger S., Roder J., Frey J., Young H.A., Ortaldo J.R.;
 RT "A cyclophilin-related protein involved in the function of natural killer cells.";
 RL Proc. Natl. Acad. Sci. U.S.A. 90:542-546(1993).
 RP REVISIONS.
 RA Anderson S.K.;
 RL Submitted (SEP-1999) to the EMBL/GenBank/DBJ databases.
 CC -1- FUNCTION: COMPONENT OF A PUTATIVE TUMOR-RECOGNITION COMPLEX.
 CC INVOLVED IN THE FUNCTION OF NK CELLS.
 CC -1- SUBCELLULAR LOCATION: Membrane-anchored. Attached to the membrane via its N-terminus.
 CC -1- SIMILARITY: Contains 1 cyclophilin-like ppiase domain.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 DR EMBL; L04288; AAA5734.2; -
 DR EMBL; AF184110; AAD56402.1; -
 DR HSSP; Q27450; 1A33
 DR Genew; HGNC:7833; NKTR.
 DR MTM; 161565; -
 DR CO; GO:0004600; F:cyclophilin; TAS.
 DR InterPro; IPR002130; CSA_Ppiase.
 DR Pfam; PR00160; pro_isomerase; 1.
 DR PRINTS; PR00153; CSA_PPIASERASE.
 DR PROSITE; PS00170; CSA_PPIASE_1; 1.
 DR PROSITE; PS0072; CSA_PPIASE_2; 1.
 KW Cyclosporin; Isomerase; Rotamase; Repeat; Membrane.
 FT DOMAIN 1 176 PPIASE, CYCLOPHILIN-TYPE.
 FT DOMAIN 219 240 ARG/LYS-RICH (BASIC).
 FT DOMAIN 421 457 ARG/LYS-RICH (BASIC).
 FT DOMAIN 970 1010 ARG/LYS-RICH (BASIC).
 FT DOMAIN 194 244 ARG/SER-RICH.
 FT DOMAIN 466 574 ARG/SER-RICH.
 FT DOMAIN 664 814 ARG/SER-RICH.
 FT DOMAIN 1311 1348 ARG-SER TANDEN REPEAT-RICH.
 SQ SEQUENCE 1462 AA; 165676 MW; D98A1147763EF527 CRC64;
 Query Match 5.4%; Score 98; DB 1; Length 1462;
 Best Local Similarity 20.0%; Pred. No. 15;
 Matches 48; Conservative 38; Mismatches 80; Indels 74; Gaps 11;
 QY 9 PAELPDRMP-----DPRPSYGAETV--VNNYIRKQGVY-----H 44
 DB 606 PVIPLSDSPPSRKQKQKPKKSTYRIQEKAKTHTLPLOSTYLANIKETGSSSSYH 665
 QY 45 RDRKQOQTEOREEMLSYGCVGWVWNSGOYPTNRLAFASFDREPKNELKNRPSRGSE 104
 DB 666 K--REKNSESPQSTYSKY-----SDRSSESPRSRSRSRS 699
 QY 105 RAEPFGRVAKSPDEKGFQAREVASVNNR--LENADESAVYDNKELANGDAL 161
 DB 700 RS-----YRSYTRRSRLASSHSRSPSSRSRKYGDHQCRRSSSYTSI 747
 QY 162 RNEDARSPFYALNTPSPFRNGNHN--DPSRKAVIYKHKHWSQDSSADPKRTKGD 219
 DB 748 SSDDGR---RAKRRLRSSGKKNSVSHKHSSESEKTLHSKYV-KGRDRSCV--RKYSE 800
 RESULT 12
 PHOE_SALTI STANDARD; PRT; 350 AA.
 ID PHOE_SALTI
 AC Q56119;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DT 15-SEP-2003 (Rel. 42, Last annotation update)
 DE Outer membrane pore protein B precursor.
 GN PHOE OR STY0365 OR T2530.
 OS Salmonella typhi.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
 CC Enterobacteriaceae; Salmonella.
 CX NCBI_Taxid=601;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=IMSS-1;
 RA Torres A., Puente J.L., Calva E.;
 RL Submitted (AUG-1993) to the EMBL/GenBank/DBJ databases.
 CC [2]
 CC SEQUENCE FROM N.A.
 RC STRAIN=CT18;
 RX MEDLINE=21534947; PubMed=11677608;

RA Parkhill J., Dougan G., James K.D., Thomson N.R., Pickard D., Main J.,
 RA Churcher C., Mungall K.L., Bentley S.D., Holden M.T.G., Sebahia M.,
 RA Baker S., Basham D., Brooks K., Chillingworth T., Connor P.,
 RA Cronin A., Davis P., Davies R.M., Dowd L., White N., Parrar J.,
 RA Feltwell T., Hamlin N., Haque A., Hien T.T., Holroyd S., Jørgels K.,
 RA Krogh A., Larsen T.S., Leather S., Moule S., O'Gaora P., Parry C.,
 RA Quail M., Rutherford K., Simmonds M., Skelton J., Stevens K.,
 RA Whitehead S., Barrett B.G.;
 RT "Complete genome sequence of a multiple drug resistant *Salmonella*
 RT enterica serovar Typhimurium CT18.";
 RL Nature 413:848-852(2001).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN-TY2 / ATCC 700931;
 RX MEDLINE=22531367; PubMed=12644504;
 RA Deng W., Liou S.-R., Plunkett G. III, Mayhew G.F., Rose D.J.,
 RA Burtland V., Kodyolani V., Schwartz D.C., Blattner F.R.;
 RT "Comparative genomics of *Salmonella enterica* serovar Typhimurium strains Ty2
 RT and CT18.";
 RL J. Bacteriol. 185:2330-2337(2003).
 CC - FUNCTION: THIS IS ONE OF THE PROTEINS INDUCED WHEN CELLS ARE
 CC GROWN UNDER PHOSPHATE LIMITATION. ITS PROTEIN PORE IS
 CC PARTICULARLY EFFICIENT IN THE UPTAKE OF INORGANIC PHOSPHATE,
 CC PHOSPHORYLATED COMPOUNDS, AND SOME OTHER NEGATIVELY CHARGED
 CC SOLUTES.
 CC - SUBUNIT: Homotrimer.
 CC - SUBCELLULAR LOCATION: Integral membrane protein. Outer membrane.
 CC - SIMILARITY: BELONGS TO THE OMPC/PHO E FAMILY OF PORENS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.ebi.ac.uk/announcements/>
 CC or send an email to license@ebi.ac.uk).
 CC -----
 DR EMBL: X74595; CAA52672.1; -
 DR EMBL: AL627266; CAD08790.1; -
 DR EMBL: AE016842; AA070114.1; -
 DR PIR: S36606; S36606.
 DR HSSP: P02932; LPHO.
 DR InterPro: IPR003429; OMP_2.
 DR InterPro: IPR001702; Porin_Gram-ve.
 DR Pfam: PF00267; Gram-ve_pore; 1.
 DR PRINTS: PR00182; ECOLI_PORIN.
 DR ProDom: PD000808; OMP_2; 1.
 DR PROSITE: PS00576; GRAM_NEG_PORIN; 1.
 KM Outer membrane; Transmembrane; Porin; Signal; Complete proteome.
 FT SIGNAL 1 20
 FT CHAIN 1 20
 FT CONFLICT 63 63 E -> K (IN REF. 1).
 FT CONFLICT 80 80 F -> G (IN REF. 1).
 FT CONFLICT 229 229 A -> G (IN REF. 1).
 SO SEQUENCE 350 AA; 38744 MW; 83055FCE5F0CC77 CRC64;

Query Match 5.38; Score 96; DB 1; Length 350;
 Best local similarity 20.58; Pred. No. 3.6;
 Matches 62; Conservative 33; Mismatches 125; Indels 82; Gaps 11;

QY 24 YGRATVNNYIRKQGVYSHRDKQKQTEKRENNISYGVAVVWNSGQVPTNRLAFA 83
 DB 34 YGRVAVK-----HYSDYDSKDG-----DQSYVAFGKKEITQIN-----DQLTGY 73
 QY 84 SPEDRFPQNELKNGRPFSGRTAFEGRAVEK--SPDEKGFQRAEVAIVNRALENAH 141
 DB 74 GRWEAFPAKNAKESDSQKTRLAFAGLKLDIGSPDYGNLGAIVDAVWDMPEFG 133
 QY 142 DESATLDNLKKELENGDALRNEDASPPYSALNRTSFKERKNGHDSRKAVITYKH 201
 DB 134 DSSAQTDNFMTRKASGAGATYRNTD-----PFGVGLDLTLQYQKNEDE-----YKQ 183
 QY 202 FWSGDRSSADKKRYGDPAPRPAFGTGLVMSRDNIPRSFTSGEGFVNFYDWGGA 261

DB 184 NGDSFGSYSD---FGSD--FAVSGAYTLSDREQLORRG-----CD 225
 QY 262 QTEADADKRWTHGHYHAPNGSIGAHVYESFNNWBGSDDRAGAYITLPSKNT 321
 DB 226 KAEE-----WATGVKYDA-----NDIYATYSSTRNW 253
 QY 322 AP 323
 DB 254 TP 255

RESULT 13
 AMPV STRCA STANDARD: PRT; 497 AA.
 ID AMPV STRCA
 AC P83053;
 DT 28-FEB-2003 (Rel. 41, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DT 15-SEP-2003 (Rel. 42, Last annotation update)
 DE Alpha-amylase, pancreatic (EC 3.2.1.1) (1,4-alpha-D-glucan
 DE glucanohydrolase) (Pancreatic alpha-amylase) (PA).
 OS *Struthio camelus* (Ostrich).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Archosauria; Aves; Palaeognathae; Struthioniformes; Struthionidae;
 OC Struthio.
 OX NCBI_TaxID=8801;
 RN [1]
 RP SEQUENCE.
 RC TISSUE=pancreas;
 RX MEDLINE=21176845; PubMed=11281265;
 RA Kabuto S., Ogawa T., Muramoto K., Oosthuizen V., Naude R.J.;
 RT "The amino acid sequence of pancreatic alpha-amylase from the ostrich,
 RT *Struthio camelus*.";
 RL Comp. Biochem. Physiol. 127B:481-490(2000).
 RN [2]
 RP SEQUENCE OF 1-53.
 RC TISSUE=pancreas;
 RA Oosthuizen V., Naude R.J., Oelofsen W., Muramoto K., Kamiya H.;
 RT "Ostrich pancreatic alpha-amylase: kinetic properties, amino terminal
 RT sequence and substrate structure.";
 RL Int. J. Biochem. 26:1313-1321(1994).
 CC - CATALYTIC ACTIVITY: Endohydrolysis of 1,4-alpha-glucosidic
 CC linkages in oligosaccharides and polysaccharides.
 CC - COFACTOR: Binds a calcium ion required for its activity. In
 CC mammals it also encloses one chloride ion which activates the
 CC enzyme.
 CC - SUBCELLULAR LOCATION: Secreted. Extracellular.
 CC - SIMILARITY: BELONGS TO FAMILY 13 OF GLYCOSYL HYDROLASES, ALSO
 CC KNOWN AS THE ALPHA-AMYLASE FAMILY.
 DR InterPro: IPR006593; Alp_amyl_cat_sud.
 DR InterPro: IPR006048; Alpha_amyl_C.
 DR InterPro: IPR006047; Alpha_amyl_cat.
 DR InterPro: IPR006046; Glyco_hydro_13.
 DR Pfam: PF00128; alpha-amylase; 1.
 DR Pfam: PF02806; alpha-amylase C; 1.
 DR PRINTS: PR00110; ALPHAMYLASE.
 DR SMART: SM00642; Amy; 1.
 DR SMART: SM00632; Amy; C; 1.
 KW Hydroxylase; Glycosidase; Carbohydrate metabolism; Calcium; Chloride;
 KM Pyridoxine carboxylic acid.
 FT MOD RES 1 1
 FT ACT_SITE 101 101
 FT ACT_SITE 195 195
 FT ACT_SITE 197 197
 FT ACT_SITE 233 233
 FT ACT_SITE 298 298
 FT ACT_SITE 299 299
 FT ACT_SITE 300 300
 FT ACT_SITE 337 337
 FT METAL 100 100
 FT METAL 159 159
 FT METAL 167 167
 FT METAL 201 201

PYRROLIDONE CARBOXYLIC ACID.
 SUBSTRATE BINDING (BY SIMILARITY).
 CHLORIDE-BINDING (BY SIMILARITY).
 BY SIMILARITY.
 CHLORIDE-BINDING (BY SIMILARITY).
 SUBSTRATE BINDING (BY SIMILARITY).
 BY SIMILARITY.
 CHLORIDE-BINDING (BY SIMILARITY).
 CALCIUM (BY SIMILARITY).
 CALCIUM (BY SIMILARITY).
 CALCIUM (BY SIMILARITY).
 CALCIUM (BY SIMILARITY).

FT DISULFID 28 86 BY SIMILARITY.
 FT DISULFID 70 115 BY SIMILARITY.
 FT DISULFID 141 160 BY SIMILARITY.
 FT DISULFID 379 385 BY SIMILARITY.
 FT DISULFID 451 463 BY SIMILARITY.
 FT CONFLICT 35 35 Y -> K (IN REF. 2).
 FT CONFLICT 50 52 IIT -> VEN (IN REF. 2).
 SQ SEQUENCE 497 AA; 55898 MW; 62DB18FCDB08CAB CRC64;
 Query Match 5.3%; Score 96; DB 1; Length 497;
 Best Local Similarity 23.0%; Pred. No. 5.6;
 Matches 57; Conservative 26; Mismatches 73; Indels 92; Gaps 13;
 QY 159 DALRNEDARSPFYSALNTPEFKENGNHDPSSMKAVYSK-HFWSGQDSSADRRKY 217
 DB 96 DAVKXMGCGSAGSGHTSGG-AVFNAGNRD---SPAVPSSGMDPNDKGTSGELIENY 151
 QY 218 GPPDAFPAPGTGLVMSKRDNIIPSPSPGSGVN--PYGFGAQIADADXTVWT-- 273
 DB 152 GDASQVRDCLVGLDLDALEKDYRSVA--GYMHLIDIGVAGFR--LDAAKHWPDP 206
 QY 274 -----HGNH-----YHANGSL- 285
 DB 207 IAAFDKLNNTWSSSGSPRPIYQVVIDGSEPIITSSQYFANRVTEFKYAKLTGI 266
 QY 286 ---GAMVYSEKERNMSGYS-----DFDR---GAVYTPFKSWNTA 322
 DB 267 RKMNEKRYA---LKNWGEQMGFVPSDRALVFNHNDNGHGAGASILTF---WDAR 319
 QY 323 PPKVQGW 330
 DB 320 LYKMAVG 327
 RESULT 14
 SKIP HUMAN
 ID SKIP HUMAN STANDARD; PRT; 536 AA.
 AC Q13573; Q13483;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Nuclear protein Skip (SKI-interacting protein) (SNM1 protein) (Nuclear
 receptor coactivator NCOA-62).
 GN SNM1 OR SKIP.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 CC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 CX NCBI_Taxid=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=98227980; PubMed=9569025;
 RA Dahl R., Wani B., Hayman M.J.;
 RT "The Ski oncoprotein interacts with Skip, the human homolog of
 RT Drosophila Bx42.";
 RL Oncogene 16:1579-1586 (1998).
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=98298162; PubMed=9632709;
 RA Baughn T.A., Kraichely D.M., Jelfcoat S.C. Jr., Winchester S.K.,
 RA Partidge N.C., Macdonald P.N.;
 RT "Isolation and characterization of a novel coactivator protein,
 RT NCOA-62, involved in vitamin D-mediated transcription.";
 RL J. Biol. Chem. 273:16434-16441 (1998).
 RN [3]
 RP SEQUENCE FROM N.A.
 RA Rowen L., Madan A., Olin S., Abbasi N., Baradaran L., Birditt B.,
 RA Bloom S., Dors M., Dickhoff R., Fleetwood P., Harrison G., James R.,
 RA Kaur A., Madan A., Owen M.P., Ratcliffe A., Shaffer T., Hood L.;
 RT "Sequencing of human chromosome 14.";
 RL Submitted (JAN-2000) to the EMBL/GenBank/DBJ databases.
 RN [4]
 RP SEQUENCE OF 282-536 FROM N.A.
 RX MEDLINE=97128797; PubMed=8973337;

RA Folk P., Puta F., Krijanova L., Blahusova A., Markos A.,
 RA Rabino M., Dotin R.P.;
 RT "The homolog of chromatin binding protein Bx42 identified in
 RT Drosophila.";
 RL Gene 181:229-231 (1996).
 CC -1- FUNCTION: INTERACTS WITH THE SKI ONCOGENE. INVOLVED IN VITAMIN D-
 CC MEDIATED TRANSCRIPTION.
 CC -1- SUBCELLULAR LOCATION: Nuclear.
 CC -1- SIMILARITY: BELONGS TO THE SNM FAMILY.
 CC -----
 CC This Swiss-Prot entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb.ch/announce/>
 CC or send an email to license@isb.ch).
 CC -----
 DR EMBL, U51432; AAC15912.1; -;
 DR EMBL, AF045184; AAC31697.1; -;
 DR EMBL, AC008372; AAR23325.1; -;
 DR EMBL, U43960; AAB48857.1; -;
 DR TRANSFAC, T04597; -;
 DR MIM, 603055; -;
 DR GO, GO:0005634; C:nucleus; TAS.
 DR GO, GO:0003713; F:transcription co-activator activity; TAS.
 DR GO, GO:0007048; P:oncogenesis; TAS.
 DR GO, GO:0006357; P:regulation of transcription from Pol II pro. .; TAS.
 DR InterPro, IPR004015; SKIP SNM.
 DR Pfam, PF02731; SKIP_SNM; I.
 KW Nuclear protein.
 FT DOMAIN 174 339 SNM.
 FT DOMAIN 219 233 PRO-RICH.
 FT DOMAIN 373 453 SH2-LIKE DOMAIN.
 SQ SEQUENCE 536 AA; 61494 MW; 0CC75E0D0B2CF842 CRC64;
 Query Match 5.3%; Score 96; DB 1; Length 536;
 Best Local Similarity 20.2%; Pred. No. 6.2;
 Matches 67; Conservative 42; Mismatches 116; Indels 106; Gaps 14;
 QY 44 HRDRKQWMBEOREWLSYGVGVWNSGY--PTN-----RLA 81
 DB 221 HSPSRKNTV-XEQQEKIPPCIS-NMKNAKGYITPDKRLADGGLQTVHINENFAKLA 288
 QY 82 PASFDEDFRNLKNGRPSSGTRAEFGVAKESPDEKGFQARREYASVW----- 133
 DB 289 EALYIADRKARE-----AVEMRAQYERMAQX--EKEKEKLEBWAQARERRAGIK 339
 QY 134 -----NRALENHDESAVLDNKKELANNDALRNEDARSPFYSAAR 175
 DB 340 THVEKDEGEAREEDRIHRDKRERQHDR-----NLSRAAPDRSKLQNNENRDISVIAL 394
 QY 176 NTPFSEKRNNGNDP---SRMKAV-----IYSKFWSGQDR-----SSSA 212
 DB 395 GVPKPRTSNVEQYDQLTFWQSGKWDSPGAGEDELYNVQDMRGKMAQSIYRPSKNL 454
 QY 213 DKRYGPPDAFPAPGTGLVMSKRDNIIPSPSPG-----EGFVNFYGFAGQTEA 265
 DB 455 DKDMYGDLEAR-----IKTNREVPDKFSGSDRRQRGRGEPVQFEDDPGLDKFL 505
 QY 266 DADKTWTHGNHHAHNSGLGMHYESKFR 296
 DB 506 EEAR--QHGGSRPDSRPKEHSEKGR 533
 RESULT 15
 IF3A MOUSE
 ID IF3A MOUSE STANDARD; PRT; 1344 AA.
 AC P23116; Q60697; Q62162;
 DT 01-NOV-1991 (Rel. 20, Created)
 DT 30-MAY-2000 (Rel. 39, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Eukaryotic translation initiation factor 3 subunit 10 (eIF-3 theta)

DE (eif3 p167) (eif3 p180) (eif3 p185) (eif3a) (p162 protein)
 DE (Centrosomin).
 GN EIF310 OR EIF3 OR CSMA.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI Taxid=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Lymphoma.
 RA Fisher R., Fillmore H., Reynolds A.B.;
 RT "Molecular cloning and characterization of the 162 kDa component
 of a multi-protein complex phosphorylated by Src."
 RT Submitted (SEP-1994) to the EMBL/Genbank/DBJ databases.
 RL [2]
 RP SEQUENCE OF 514-961 FROM N.A. (CENTROSOMIN B).
 RA MEDLINE=98039715; PubMed=9372446;
 RA Petzelt C., Joswig G., Mincheva A., Lichter P., Stammer H., Werner D.;
 RT "The centrosomal protein centrosomin A and the nuclear protein
 centrosomin B derive from one gene by post-transcriptional processes
 involving RNA editing."
 RT J. Cell Sci. 110:2573-2578 (1997).
 RN [3]
 RP SEQUENCE OF 514-790 FROM N.A. (CENTROSOMIN A).
 RA MEDLINE=91277032; PubMed=1829089;
 RA Joswig G., Petzelt C., Werner D.;
 RT "Murine cDNAs coding for the centrosomal antigen centrosomin A."
 RL J. Cell Sci. 98:37-43 (1991).
 RN [4]
 RP REVISIONS.
 RA Joswig G., Petzelt C., Werner D.;
 RL Submitted (DEC-1996) to the EMBL/Genbank/DBJ databases.
 CC -!- FUNCTION: BINDS TO THE 40S RIBOSOME AND PROMOTES THE BINDING OF
 METHIONYL-TRNAI AND MRNA.
 CC -!- SUBUNIT: EIF-3 IS COMPOSED OF AT LEAST 12 DIFFERENT SUBUNITS.
 CC -!- SUBCELLULAR LOCATION: Cytoplasmic (by similarity).
 CC -!- DOMAIN: CONTAINS 1 PCI DOMAIN.
 CC -!- PTM: PHOSPHORYLATED.
 CC -!- SIMILARITY: BELONGS TO THE EIF310 FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: U14172; AA090910.1; -;
 DR EMBL: X84651; CA059144.1; -;
 DR EMBL: X17373; CA05246.1; -;
 DR PIR: T42637; T42637.
 DR MGD: MGI:95301; Eif3.
 DR InterPro: IPR000717; PCI.
 DR Pfam: PF01399; PCI.1.
 DR SMART: SM00088; PINT.1.
 KM Initiation factor; Protein biosynthesis; Repeat; Phosphorylation.
 FT DOMAIN 924 1133
 FT 21 X 10 AA TANDDEM REPEAT OF D-[DS]-D-R-
 FT [GP]-[PS]-[RW]-R-[GN]-[AM].
 FT RQAKREKERKIQEHQIKKIKTVREERLEQIKTE ->
 FT PRGKAREGRTSRTRRANQENSGAVRADQDR (IN
 FT REF. 2 AND 3).
 FT CONFLICT 613 647
 FT EL -> DY (IN REF. 2 AND 3).
 FT CONFLICT 717 717
 FT Q -> H (IN REF. 2 AND 3).
 FT CONFLICT 766 766
 FT A -> V (IN REF. 2 AND 3).
 FT CONFLICT 787 790
 FT RHR -> SIVA (IN REF. 3).
 FT CONFLICT 793 793
 FT E -> D (IN REF. 2).
 SQ SEQUENCE 1344 AA; 161949 MW; F4CAR2169F57712 CRC64;

Query Match 5.3%; Score 96; DB 1; Length 1344;
 Best Local Similarity 18.9%; Pred. No. 20;
 Matches 53; Conservative 38; Mismatches 103; Indels 86; Gaps 11;

QY 3 DDRVTPAPBLRMDPYRPSYGRATVNNYIRK--WQYYSRDR-----KQMTBEQ 56
 DB 1115 DDRV--PRGGDARPGPRP-----FKKGWREKREKARESGWGPRESRPSSE 1161
 QY 57 RHLSTGCGVGVWVNSGYPTNRLAFASPFEDRFKELKNGRPPSGETRAEPFGRIYAKS 116
 DB 1162 RHW-----DRDEKRDNDQREENDKLEDRDRREED 1193
 QY 117 FDEEKGFORAREVATVNNRLAENADEAYLDNLKKELANGNDALRNEDARSPYSALRN 176
 DB 1194 GDRDRRRRRRDEGGRGPR---AESSSWDSRRDDDRDRRRDRDR-----RD 1243
 QY 177 TPEFKERNGNHPSRMKAVIYKH-----FWSGQDRSSADKKYGPDAFRPAPGTGLV 232
 DB 1244 LRLDRRLDRDRDRRGPPRLSRDEEASGWRTD--DRDDRTERRDPRRVPPA----- 1298
 QY 233 DMSDRNIPSPSPSGSGVFVFGMGAGTADALDKITW 272
 DB 1299 -LSRDRRRR-----EREGEKAKSW 1318

Search completed: September 22, 2003, 12:44:17
 Job time : 18 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 22, 2003, 12:42:00 ; Search time 40 Seconds

(without alignments)
2135.385 Million cell updates/sec

Title: US-09-884-948-1

Perfect score: 1811
Sequence: 1 DSDRVTPPAEPJDRMPDY.....ITFIPKSWNTAPDKVKQGP 331

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 830525 seqs, 258052604 residues

Total number of hits satisfying chosen parameters: 830525

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database :

1: SP_archaea:*
2: SP_bacteria:*
3: SP_fungi:*
4: SP_human:*
5: SP_invertebrate:*
6: SP_mammal:*
7: SP_mirc:*
8: SP_organelle:*
9: SP_phage:*
10: SP_plant:*
11: SP_rodent:*
12: SP_virus:*
13: SP_vertebrate:*
14: SP_unclassified:*
15: SP_virus:*
16: SP_bacteriaph:*
17: SP_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1811	100.0	376	2	Q9ZAF5 streptomyc
2	1811	100.0	407	2	Q8KRJ2 streptomyc
3	1732	95.6	334	2	Q8KNY5 streptomyc
4	1487	82.1	416	2	Q8GR90 streptomyc
5	1432	79.1	411	2	Q9RIS1 streptomyc
6	1432	6.7	737	12	Q8Q0G0 adeno-assoc
7	114.5	6.3	1217	5	Q17240 bombay mori
8	113.5	6.3	738	12	Q8Q0F8 adeno-assoc
9	112.5	6.2	1161	5	Q9W2X8 adeno-assoc
10	111.5	6.2	758	10	Q8I5Z5 oryza sativ
11	110	6.1	736	12	Q65311 adeno-assoc
12	109	6.0	919	2	Q07351 adeno-assoc
13	108	6.0	401	4	Q96C72 vibrio chol
14	108	6.0	558	4	Q9H0G5 homo sapien
15	107	5.9	736	12	Q56139 adeno-assoc
16	106.5	5.9	736	12	Q56137 adeno-assoc

17	106.5	5.9	736	12	Q9W2E8	Q9W2E8 adeno-assoc
18	106.5	5.9	894	16	Q8PB66	Q8PB66 xanthomonas
19	106.5	5.9	918	16	Q9KXU6	Q9KXU6 vibrio chol
20	106	5.9	776	5	Q8STU8	Q8STU8 encephalit
21	105.5	5.8	839	11	Q8C235	Q8C235 mus musculu
22	105.5	5.8	865	3	Q8X0P4	Q8X0P4 neurospora
23	105.5	5.8	1888	11	Q88466	Q88466 mus musculu
24	105	5.8	325	3	P78890	P78890 schizosacch
25	105	5.8	920	10	Q9SLT4	Q9SLT4 arabidopsis
26	105	5.8	921	10	Q9SLT3	Q9SLT3 arabidopsis
27	105	5.8	1092	10	Q9LO55	Q9LO55 arabidopsis
28	104	5.7	219	16	Q8W0Z8	Q8W0Z8 staphylococ
29	104	5.7	636	10	Q8S252	Q8S252 oryza sativ
30	103.5	5.7	1245	11	Q8R314	Q8R314 mus musculu
31	102.5	5.7	12	12	Q9PZV7	Q9PZV7 hepatitis d
32	102	5.6	219	16	Q99RJ9	Q99RJ9 staphylococ
33	102	5.6	504	5	Q16352	Q16352 caenorhabdi
34	102	5.6	735	12	Q56652	Q56652 adeno-assoc
35	101.5	5.6	383	16	Q8YF76	Q8YF76 brucella me
36	101.5	5.6	383	16	Q8G2P2	Q8G2P2 brucella me
37	100.5	5.5	285	5	Q8WRP0	Q8WRP0 caenorhabdi
38	100.5	5.5	467	10	Q94LH8	Q94LH8 oryza sativ
39	100.5	5.5	801	11	Q8B022	Q8B022 mus musculu
40	100.5	5.5	966	11	Q8CB26	Q8CB26 mus musculu
41	100	5.5	630	5	Q9GNN5	Q9GNN5 branchiosto
42	100	5.5	730	5	Q8T867	Q8T867 mus musculu
43	100	5.5	744	11	Q8BR43	Q8BR43 mus musculu
44	100	5.5	760	11	Q8BNM7	Q8BNM7 mus musculu
45	100	5.5	775	5	Q9Y015	Q9Y015 plasmodium

ALIGNMENTS

RESULT 1

ID Q9ZAF5 PRELIMINARY; PRT; 376 AA.
AC Q9ZAF5;
DT 01-MAY-1999 (TREMBLrel. 10, Created)
DT 01-MAY-1999 (TREMBLrel. 10, Last sequence update)
DT 01-MAR-2002 (TREMBLrel. 20, Last annotation update)
DE Transglutaminase (EC 2.3.2.13) (Fragment).
OS Streptomyces mobaraensis.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Streptomyces; Streptomycetaceae; Streptomyces.
CX NCBI_TaxID=35621;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=DSMZ;
RX MEDLINE=99053680; PubMed=9839945;
RA Pasternack R., Dorsch S., Ottentach J.T., Robenek I.R., Wolf S.,
RT Fuchsbauer H.L.;
RT "Bacterial pro-transglutaminase from Streptomyces mobaraensis :
RT purification, characterisation and sequence of the zymogen.";
RL Eur. J. Biochem. 257:570-576(1998).
DR EMBL; Y18315; CAA7128.1; -;
KW Acyltransferase; Transferase.
FT NON TER
FT SEQUENCE 1
SQ SEQUENCE 376 AA; 42445 MW; 15FE7474DE3771B9 CRC64;

Query Match 100.0%; Score 1811; DB 2; Length 376;
Best Local Similarity 100.0%; Pred. No. 1,4e-128;
Matches 331; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	1	DSDRVTPPAEPJDRMPDYPRSYGAETVVNNYTRKMQVYSHRGRQKMTESREWL	60
DB	46	DSDRVTPPAEPJDRMPDYPRSYGAETVVNNYTRKMQVYSHRGRQKMTESREWL	105
QY	61	SYGCVGVTVWNSGQYPTNRIAFASFEDEDRFKNELKNGRPPSGETRAEFGRVAKESFDEE	120
DB	106	SYGCVGVTVWNSGQYPTNRIAFASFEDEDRFKNELKNGRPPSGETRAEFGRVAKESFDEE	165
QY	121	KGFQARAEVASVMNRALENAHDSAYLIDNKKELANGDNLINEDARSPFYALNTPSF	180

```

Db      166 KGFQARREAVASVMMRALENADHESAYLDNLKKELANGDARNEDARSPFYGALNTPSF 225
QY      181 KERNGNHNDPSRMKAIVYSKHEWSGODRSSADKKRYGDDPAPRPAFGVLDMGRDNI 240
Db      226 KERNGNHNDPSRMKAIVYSKHEWSGODRSSADKKRYGDDPAPRPAFGVLDMGRDNI 285
QY      241 PRSPTSPGEGFVNFYGMFGAQTADADACTVTHGNHYHAPNGSLGAMHYESKFRNWE 300
Db      286 PRSPTSPGEGFVNFYGMFGAQTADADACTVTHGNHYHAPNGSLGAMHYESKFRNWE 345
QY      301 GYSDFDRGAYVITFIPKSNMTPADKYKQGW 331
Db      346 GYSDFDRGAYVITFIPKSNMTPADKYKQGW 376

RESULT 2
Q8KRJ2 ID Q8KRJ2 PRELIMINARY; PRT; 407 AA.
AC Q8KRJ2;
DT 01-OCT-2002 (TrEMBLrel. 22, Created)
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE Transglutaminase precursor.
OS Streptomyces mobaraensis.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Streptomycinae; Streptomycetaceae; Streptomyces.
OX NCBI_TaxID=35621;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=IFO13819;
RA Kikuchi Y., Date M., Yokoyama K.-I., Umezawa Y., Matsui H.;
RT "Secretion of active form transglutaminase of Streptocorticillium
RT mobaraense in Corynebacterium glutamicum: Processing of pro-domain
RT with co-secreted subtilisin-like protease from Streptomyces
RT aliostrictolus.";
RL Submitted (JUL-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF531437; AAM9595.1; -.
KW Signal.
FT SIGNAL.
FT CHAIN.
FT SIGNAL.
SQ SEQUENCE 407 AA, 45664 MW, 107777A04EAB2DF4 CRC64;

Query Match 100.0%; Score 1811; DB 2; Length 407;
Best Local Similarity 100.0%; Pred. No. 1.6e-128;
Matches 331; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DSDDRVTPPAEPLDMPPRPSYGRAEVYNNYIRKQYVYSHRDRKQOMTEQRRL 60
Db 77 DSDDRVTPPAEPLDMPPRPSYGRAEVYNNYIRKQYVYSHRDRKQOMTEQRRL 136
QY 61 SYGCVGVTVWNSGQYFPNRLAFASFDDEDFKXELKNGRPSGSEFRAEFGRVAKESFDEE 120
Db 137 SYGCVGVTVWNSGQYFPNRLAFASFDDEDFKXELKNGRPSGSEFRAEFGRVAKESFDEE 196
QY 121 KGFQARREAVASVMMRALENADHESAYLDNLKKELANGDARNEDARSPFYGALNTPSF 180
Db 197 KGFQARREAVASVMMRALENADHESAYLDNLKKELANGDARNEDARSPFYGALNTPSF 256
QY 181 KERNGNHNDPSRMKAIVYSKHEWSGODRSSADKKRYGDDPAPRPAFGVLDMGRDNI 240
Db 257 KERNGNHNDPSRMKAIVYSKHEWSGODRSSADKKRYGDDPAPRPAFGVLDMGRDNI 316
QY 241 PRSPTSPGEGFVNFYGMFGAQTADADACTVTHGNHYHAPNGSLGAMHYESKFRNWE 300
Db 317 PRSPTSPGEGFVNFYGMFGAQTADADACTVTHGNHYHAPNGSLGAMHYESKFRNWE 376
QY 301 GYSDFDRGAYVITFIPKSNMTPADKYKQGW 331
Db 377 GYSDFDRGAYVITFIPKSNMTPADKYKQGW 407

RESULT 3

```

```

Q8KRY5 ID Q8KRY5 PRELIMINARY; PRT; 334 AA.
AC Q8KRY5;
DT 01-OCT-2002 (TrEMBLrel. 22, Created)
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE Transglutaminase (Fragment).
OS Streptomyces mobaraensis.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Streptomycinae; Streptomycetaceae; Streptomyces.
OX NCBI_TaxID=35621;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 27441;
RA Jiang S.-T., Tzeng S.-S., Wu W.-T., Chen G.-H.;
RL Submitted (JUL-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AY129279; AAN0353.1; -.
FT NON_TER.
FT SIGNAL.
SQ SEQUENCE 334 AA, 38168 MW, 8C2356C46FC2CB35 CRC64;

Query Match 95.6%; Score 1732; DB 2; Length 334;
Best Local Similarity 93.7%; Pred. No. 1.1e-122;
Matches 310; Conservative 14; Mismatches 7; Indels 0; Gaps 0;

QY 1 DSDDRVTPPAEPLDMPPRPSYGRAEVYNNYIRKQYVYSHRDRKQOMTEQRRL 60
Db 4 DSDDRVTPPAEPLDMPPRPSYGRAEVYNNYIRKQYVYSHRDRKQOMTEQRRL 63
QY 61 SYGCVGVTVWNSGQYFPNRLAFASFDDEDFKXELKNGRPSGSEFRAEFGRVAKESFDEE 120
Db 64 SYGCVGVTVWNSGQYFPNRLAFASFDDEDFKXELKNGRPSGSEFRAEFGRVAKESFDEE 123
QY 121 KGFQARREAVASVMMRALENADHESAYLDNLKKELANGDARNEDARSPFYGALNTPSF 180
Db 124 KGFQARREAVASVMMRALENADHESAYLDNLKKELANGDARNEDARSPFYGALNTPSF 183
QY 181 KERNGNHNDPSRMKAIVYSKHEWSGODRSSADKKRYGDDPAPRPAFGVLDMGRDNI 240
Db 184 KERNGNHNDPSRMKAIVYSKHEWSGODRSSADKKRYGDDPAPRPAFGVLDMGRDNI 243
QY 241 PRSPTSPGEGFVNFYGMFGAQTADADACTVTHGNHYHAPNGSLGAMHYESKFRNWE 300
Db 244 PRSPTSPGEGFVNFYGMFGAQTADADACTVTHGNHYHAPNGSLGAMHYESKFRNWE 303
QY 301 GYSDFDRGAYVITFIPKSNMTPADKYKQGW 331
Db 304 GYSDFDRGAYVITFIPKSNMTPADKYKQGW 334

RESULT 4
Q8GR90 ID Q8GR90 PRELIMINARY; PRT; 416 AA.
AC Q8GR90;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Transglutaminase.
OS Streptomyces cinamomeus.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Streptomycinae; Streptomycetaceae; Streptomyces.
OX NCBI_TaxID=33446;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=IFO 12852;
RA Yokoyama K.;
RT "Streptocorticillium cinamomeum IFO12852 Tgase gene.";
RL Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB085638; BAC24766.1; -.
SQ SEQUENCE 416 AA, 46394 MW, 677DA9778829839D CRC64;

Query Match 82.1%; Score 1487; DB 2; Length 416;
Best Local Similarity 81.5%; Pred. No. 4.4e-104;

```

Matches 263; Conservative 26; Mismatches 35; Indels 0; Gaps 0;

QY 2 SDDRVTTPAEPDMPDPYRPSYGRATVNNYIRKQOYSHRDGRKQOITEOREWLS 61
 DB 87 SDDRTPPAEPDMPDMPAAYAGRAITVNNYIRKQOYSHRDGRKQOITEOREWLS 146
 QY 62 YGCVGYTVWNSGQPTNRLAFASFDPRPKMLKXGRPSGGEFRAEFGRVAKESFDEEK 121
 DB 147 YGCVGYTVWNSGQPTNRLAFASFDPRPKMLKXGRPSGGEFRAEFGRVAKESFDEEK 206
 QY 122 GFORAREVAVNNALNADHESAYLDNKKELANGDARNEDASPPYSALRTPSEFK 181
 DB 207 GFORARDVAVNNALNADHESAYLDNKKELANGDARNEDASPPYSALRTPSEFK 266
 QY 182 BENGANDPSSMKAVIYSKHFWSGQDRSSADKRYGDPAPFAPAGTGLVMSDRNP 241
 DB 267 BNDGNDPSSMKAVIYSKHFWSGQDRSSADKRYGDPAPFAPAGTGLVMSDRNP 326
 QY 242 RSPFSPGEGFVNDYGFQATADADKTWTHGNHHAHNSGLAMHYESKPRNMSG 301
 DB 327 RSPAPGEGVNFYGFQATADADKTWTHGNHHAHNSGLAMHYESKPRNMSG 386
 QY 302 YSDFPRGAYVITFIPKSNMTAPDKYQGW 331
 DB 387 YADFPRGAYVITFIPKSNMTAPDKYQGW 416

RESULT 5

Q9RIS1 PRELIMINARY; PRT; 411 AA.
 ID Q9RIS1; 01-MAY-2000 (Tremblrel. 13, Created)
 DT 01-MAY-2000 (Tremblrel. 13, Last sequence update)
 DT 01-MAR-2002 (Tremblrel. 20, Last annotation update)
 DS Transglutaminase precursor (EC 2.3.2.13).
 GN TABSE
 OS Streptomyces cinamomeus.
 OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
 OC Streptomyces; Streptomycetaceae; Streptomyces.
 OX NCBI_TaxID=53446;
 RN 11
 RP SEQUENCE FROM N.A.
 RC SMRAIN=CHS 683.68;
 EX MEDLINE=98336622; PubMed=9672751;
 RA Duran R., Juncua M., Schmitter J.M., Gancet C., Goulas P.,
 RT "Purification, characterization, and gene cloning of transglutaminase
 RT (Tgase) from Streptococcus cinamomeus CHS 683.68".
 RL Biochimie 80:313-319(1998).
 DR EMBL: Y08820; CAA70055.1;
 KM Acyltransferase; Signal; Transferase.
 FT SIGNAL 81
 FT CHAIN 82 411
 SQ SEQUENCE 411 AA; 46499 MW; 01A0C2A2EF4C388B CRC64;

Query Match 79.1%; Score 1432; DB 2; Length 411;
 Best Local Similarity 79.7%; Pred. No. 6, 1e-100;
 Matches 263; Conservative 27; Mismatches 36; Indels 4; Gaps 3;

QY 2 SDDRVTTPAEPDMPDPYRPSYGRATVNNYIRKQOYSHRDGRKQOITEOREWLS 61
 DB 86 SDDRTPPAEPDMPDMPAAYAGRAITVNNYIRKQOYSHRDGRKQOITEOREWLS 145
 QY 62 YGCVGYTVWNSGQPTNRLAFASFDPRPKMLKXGRPSGGEFRAEFGRVAKESFDEEK 121
 DB 146 YGCVGYTVWNSGQPTNRLAFASFDPRPKMLKXGRPSGGEFRAEFGRVAKESFDEEK 204
 QY 122 GFORAREVAVNNALNADHESAYLDNKKELANGDARNEDASPPYSALRTPSEFK 181
 DB 205 GFORARDVAVNNALNADHESAYLDNKKELANGDARNEDASPPYSALRTPSEFK 263
 QY 182 BENGANDPSSMKAVIYSKHFWSGQDRSSADKRYGDPAPFAPAGTGLVMSDRNP 241
 DB 264 BNDGNDPSSMKAVIYSKHFWSGQDRSSADKRYGDPAPFAPAGTGLVMSDRNP 323

QY 242 RSPFSPGEGFVNDYGFQATADADKTWTHGNHHAHNSGLAMHYESKPRNMSG 301
 DB 324 RSPAPGEGVNFYGFQATADADKTWTHGNHHAHNSGLAMHYESKPRNMSG 383
 QY 302 YSDFPRGAYVITFIPKSNMTAPDKYQGW 331
 DB 384 YADFPRGAYVITFIPKSNMTAPDKYQGW 411

RESULT 6

Q8UQ0 PRELIMINARY; PRT; 737 AA.
 ID Q8UQ0; 01-OCT-2002 (Tremblrel. 22, Created)
 DT 01-OCT-2002 (Tremblrel. 22, Last sequence update)
 DT 01-MAR-2003 (Tremblrel. 23, Last annotation update)
 DS Capsid protein.
 OS Adeno-associated virus 7.
 OC Viruses; ssDNA viruses; Parvoviridae; Parvovirinae; Dependovirus.
 OX NCBI_TaxID=202812;
 RN 11
 RP SEQUENCE FROM N.A.
 RX PubMed=12192090;
 RA Gao G.P., Alviria M.R., Wang L., Calcedo R., Johnston J., Wilson J.M.;
 RT "Novel adeno-associated viruses from rhesus monkeys as vectors for
 RT human gene therapy".
 RL Proc. Natl. Acad. Sci. U.S.A. 99:11854-11859(2002).
 DR EMBL: AF513851; AA03855.1;
 DR InterPro: IPR001403; Parvo_coat.
 DR Pfam: PF00740; Parvo_coat; 1.
 SQ SEQUENCE 737 AA; 81652 MW; D63BCE7583140D0 CRC64;

Query Match 6.7%; Score 122; DB 12; Length 737;
 Best Local Similarity 22.8%; Pred. No. 0.81;
 Matches 76; Conservative 45; Mismatches 161; Indels 52; Gaps 16;

QY 18 DYPSPYGRATVNNYIRKQOYSHRDGRKQOITEOREWLSYGVGYTVWNSGQY-- 75
 DB 4 DGYLPDW--LEDNLSRGIREW---WDLKPGAHPKXANQKQDNGGLVLPQYKYLGFENG 58
 QY 76 -----PNNRLAFSPEDR-FNLEKNG-RP--RSGETAEFEFGVAESDEKGRQRA 126
 DB 59 LDKGEPVNAADALHEDPAVDQKAGDNNYLYNHADBFQRLQ---DTSFGANLG 115
 QY 127 REVASVNNRLENADHESAYLDNKKELANGDARNEDASPPYSALRTPSEFKRNG 186
 DB 116 RAVFOAKKRVLEPL-----GLVESGAKTAPAKKRPVBPQSPSSGIGK---- 163
 QY 187 NHDSEKKAIVISKHFWSGQDRSSADKRYGDPAPFAPAGTGLVMS-----RQKNIP 241
 DB 164 GQOQAPRRRL-----NGQGTGDESVPOPLGEPAPAPSVSGTVAAAGGAPVADNNG 218
 QY 242 RSPFSPGEGFVNDYGFQATADADKTW--TGNNYHAHNSGLAMHYESKPRNMSG 298
 DB 219 ADGYNASGWHDDSTWLDRTITSTRT-WALPTNNHLYQIISSETAGSTNDNTITFY 277
 QY 299 SE--GYSDPRGAYVITFIPKSNMTAPDKYQGW 330
 DB 278 STPWGYDFNRR--FCHGFSRDMQRL--INNWN 306

RESULT 7

Q17240 PRELIMINARY; PRT; 1217 AA.
 ID Q17240; 01-NOV-1996 (Tremblrel. 01, Created)
 DT 01-NOV-1996 (Tremblrel. 01, Last sequence update)
 DT 01-OCT-2002 (Tremblrel. 22, Last annotation update)
 DS SericinB.
 OS Bombyx mori (Silk moth).
 OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
 OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
 OC Neoptera; Endopterygota; Lepidoptera; Glossata; Ditrysia; Bombycoidea;

```

OC Bombycidae, Bombyx.
OX NCBI_TaxID=7091;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=European 200x300; TISSUE=Middle silk gland;
EX MEDLINE=97362906; PubMed=9219370;
RA Garel A.A., Deleage G.G., Prudhomme J.J.;
RT "Structure and organization of the Bombyx mori Sericin 1 gene and of
  the Sericin1 deduced from the sequence of the Serib cDNA.";
RL Insect Biochem. Mol. Biol. 27:469-477(1997).
DR EMBL: Z48802; CA88741.1;
SQ SEQUENCE 1217 AA; 123411 MW; 4C5789F87F8666D0 CRC64;

Query Match 6.3%; Score 114.5; DB 5; Length 1217;
Best Local Similarity 20.4%; Pred. No. 5.8;
Matches 60; Conservative 47; Mismatches 136; Indels 51; Gaps 11;

QY 33 NTRKQOYVSHRDKQKQTEBQ---EMLSYGVGVTVVNSQ-----YPTNR 79
DB 124 NVVSDQAVAAASDABRENRSADQNNQANWMDGSGVYADRSASRRRQANYSDKD 183
QY 80 LAFSPEDRFENKELKGRPRSGEPTAEFEGRVAKESPEEKQFORARVASYMRLALEN 139
DB 184 ITPASKDSRADSSRSN-----AYNRDSD-----GSBSAGLSDBSAS 223
QY 140 AHDESAIYDNLKKELANGNDALRNEDARSPFYALRNTPSFEKRNNGNDPS----RMK 194
DB 224 SKNDNFVYRTKDSI--GGQAKSSRSBSHSGESDAYNNSPDGYNAGTRSSITNKKKAS 281
QY 195 AVTYS--KIFMSGGDRSSADKRYGDPDAFPAPAGTGLVMSMDRNPSPSPGEGFV 252
DB 282 STYADKQDIRAANDRSSSKOLKSSAQSISGP-KATSV--SSKDRQYNSDKKSKSDAYV 338
QY 253 NPFYGMFG--AQTEADADKTYWTHGCHYHAPGSLGAAHYVESKFRWMSGYSD 304
DB 339 GRD---GTVATSNKDSKTSRQSNITVADQNSVSDASADQTSKYDRGISD 388

RESULT 8
Q8JCF8 PRELIMINARY; PRT; 738 AA.
AC Q8JCF8;
DT 01-OCT-2002 (TREMBLrel. 22, Created)
DT 01-OCT-2002 (TREMBLrel. 22, Last sequence update)
DT 01-MAR-2003 (TREMBLrel. 23, Last annotation update)
DE Capsid protein.
OS Adeno-associated virus 8.
OC Viruses; ssDNA viruses; Parvoviridae; Parvovirinae; Dependovirus.
OX NCBI_TaxID=202813;
RN [1]
RP SEQUENCE FROM N.A.
RX PubMed=12192090;
RA Gao G.P., Alivira M.R., Wang L., Calcedo R., Johnston J., Wilson J.M.;
RT "Novel adeno-associated viruses from rhesus monkeys as vectors for
  human gene therapy.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:11854-11859(2002).
DR EMBL: AF513852; JAM03857.1;
DR InterPro: IPR001403; Parvoviridae.
DR Pfam: PF00740; Parvoviridae.
SQ SEQUENCE 738 AA; 81756 MW; 59BC2BE1B148AA14 CRC64;

Query Match 6.3%; Score 113.5; DB 12; Length 738;
Best Local Similarity 22.9%; Pred. No. 3.5;
Matches 78; Conservative 46; Mismatches 153; Indels 63; Gaps 20;

QY 18 DVPSPSGAETVANNYTRKQOYVSHRDKQKQTEBQREMLSYGVGVTVVNSQY-- 75
DB 4 DGYLPDW--LEDNSEGRKEM--WALKPAPAPKPAHQKQDDRGGLVDPGYRLSPFPG 58
QY 76 -----PTNRLAFASFDKDR-FKMKLKG-RP-RSGETRAEFGRVAKESFDEKQFORA 126
DB 59 LDKGEPVNNADAAALAHDKAVDQQLQAGNDPFLRVNVAHDAFQGRLE--DISFGNLG 115

```

```

QY 127 REVASVNNRALENADHESAVLNLKKELANGNDALRNEDARSPFYALRNTPSFEKRN 186
DB 116 RAVFOAKRVLLEPL-----GLVEGAKTAPKPKPVPSPQSPDSSTQIGKK---- 163
QY 187 NNDPBRMAAVYVSHFMSGODSSADKRYGDPDARAP-CTGLVMSMDRNPSPSPGEGFV 245
DB 164 GOQPARKRL-----NFGQTGDSVDPDQPGEEPPA--APSGVGNTPAAQGGAPMADN 215
QY 246 SPG-----EGFVNFQYGMFGAQTEADADKTYW--THQNHYH--APNGSIGAM--HV 290
DB 216 NRGADGVSSSSGMMHCDSTGLDRTVTTSTRT-WALFTYNNHLYKQISNGTGGATINDT 274
QY 291 YSKKFRNWSBGYSDFRGAVITFIPKSMNTAPDKVQGW 330
DB 275 YFGYSTPW--GYDPFNR--FHCHFSRDMQRL--INNNV 307

RESULT 9
Q9W2X8 PRELIMINARY; PRT; 1161 AA.
AC Q9W2X8;
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DT 01-OCT-2002 (TREMBLrel. 22, Last annotation update)
DE CG15311 protein.
OS Drosophila melanogaster (fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
  Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
  Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
  Amanatides P.G., Scherer S.E., Li P.W., Hostkins R.A., Galle R.F.,
  George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
  Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
  Brandon R.C., Rogers Y.-H.C., Blazer R.G., Champagne M., Pfeiffer B.D.,
  Man K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
  Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
  Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
  Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
  Borokova D., Botchan M.R., Bouck J., Brokstein P., Brotter P.,
  Burkova D.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
  Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
  de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
  Dodson K., Doug L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
  Durbin K.J., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W.,
  Foster C., Gabrielian A.E., Gary N.S., Gelbart W.M., Glasser K.,
  Glodok A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
  Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
  Hostin D., Houston K.A., Howland T.J., Wei M.-H., Idagawa C.,
  Kamei M., Kalush F., Karpen G.H., Ke Z., Kentonson U.A., Ketchum K.A.,
  Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kuip D., Lai Z.,
  Liao P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
  Liu X., Mattet B., McIntosh T.C., McLeod M.P., McPherson D.,
  Merkulov G., Mishina N.V., Mobarry C., Morris J., Moshrefi A.,
  Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
  Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Paclad J.M.,
  Palazolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
  Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
  Shie B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
  Spier E.G., Spradling A.C., Stapleton M., Strong R., Sun E.,
  Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
  Wang Z.-Y., Wasman D.A., Weinstein G.M., Weissbach J.,
  Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
  Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
  Zhang X.H., Zhong P.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
  RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
  RT "The genome sequence of Drosophila melanogaster.";
  RL Science 287:2185-2195(2000).

```

DR EMBL; AE003449; AAF46559.1; -.
DR FlyBase; FBgn030182; CG15311.
SQ SEQUENCE 1161 AA; 130214 MW; DD01D5B408051D36 CRC64;

QY 8 PRAEFLDRNDPYPSPSYGRASFTVNNYIRKMQOVSHPDGRKQWTEBOEWLSTGCVGV 67
| : : : : :
D5 180 POLEREDVEHEENANFGF-QSFAYKKLKIMHETQSQQDKRGGDDGE----- 226

```
DY      68 TWANVGQYPNRLTAFSFDIEDRFNELNLNGRP---RGGETRAFEFGR-----VAKESFDE 119
          : |::| ::|   ||     ||| ::| ::| ::| ::| ::| ::| ::| ::| ::| ::| ::| ::| ::| ::| ::| ::|
Dz      227 VDLLEDGEYPPSQL----FDAIEITNERKFNKPASQGSGTSRPOLOKKDKDDQMKQEQVO 282
```

DQ
120 EKGFRRAREVAVNMRRALNAHDESAVDNLKELANNDALNRBDARSPEVSALRNTPS 179

DB
283 EQGHRKSRYDVAIGDIENLDNDELPLPTNNGEDEDGGDDADIDDD--EDIKSALDNDEL 340

180 FKKKGGNHDPSSKKKAVIYSKHFWSGGQSSSSADKKKXGDPDFKFAFGTLDVMSRDKN 239
 341 AKKYPVATSTTTKVPPTTLATSKTTSRSSSSSTTTTMASTTATSSPSSTTTK----- 394

Db

```
240  LKSLISFSDG-----FVNDIGMGAQLADDAALVMIGHNIAEEN 262  
      ||| | | | : :: | | | |  
395 -PRFT-PTGRKLRRNFGLSQPETRYGSYDIGRIRAQNGED-DT-----EEKD 444
```

Db 445 ENVEGHHYDTGGSSSRKLVSFDEPKSEENIYTIYFGKON-ATEKKQC 493

RESULT 10
Q8L525
ID Q8L525
PRELIMINARY:
PRT, 758 AA.

DT	01-OCT-2002 (TREMblrel, 22, Created)
DT	01-OCT-2002 (TREMblrel, 22, last sequence update)
DT	01-MAR-2003 (TREMblrel, 23, last annotation update)

GN B1103C09.11 OR P045ID05.19.
OS *Oryza sativa* (japonica cultivar-group).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
CC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae.

CC	Ehrhartoideae; Oryzeae; Oryza.
OX	NCBI_TaxID=39947;
RN	[1]
RP	SEQUENCE FROM N.A.

RC STRAIN=cv. Nipponbare;
RA Sasaki T., Yamamoto K.;
RT "Oryza sativa (japonica cultivar-group) genomic DNA, chromosome 1, BAC clone:B1103C09";

RL submitted (FEB-2001) to the EMBL/Genbank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 EC STRAIN=cv. Nipponbare;

RA sasakii 1., Matsubara 1., Idemitsu K.?
RT "Oryza sativa (japonica cultivar-group) genomic DNA, chromosome 1, PAC
RT clone:P0451D05."
RL Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.

DR EMBL; AF003253; BAB92317.1; -.
DR Gramene; Q8L525; -.
SQ SEQUENCE 758 AA; 81761 MW; 2A95E763198FE12C CRC64;

Query Match	6.2%;	Score 11.5;	DB 10;	Length 758;
Best Local Similarity	22.0%;	Pred. No. 5.2;		
Matches	78;	Conservative	40;	Mismatches 146;
			Indels	91;
			Gaps	14

```

2Y      3 DDRVPPPAEPLDRMDPYRPSYGRAEVTV--NNYIRKMQ-----YTSHRD-----46
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      125 DGRAPBPKEVAGAGQGDGEKDVGEAAVEGDGHEKKQEPFAVPTSGAFYVHDDRFQE 184

```

```

RESULT 11
Q65311
ID Q65311 PRELIMINARY; PRT; 736 AA.
AC Q65311;
DT 01-NOV-1996 (TReMBLrel. 01, Created)
DT 01-NOV-1996 (TReMBLrel. 01, Last sequence update)
DT 01-JUN-2001 (TReMBLrel. 17, Last annotation update)
DE Capsid protein.
DE Agemo-associated virus 3.
OS Viruses; ssDNA; Viruses; Parvoviridae; Parvovirinae; Dependovirus.
OC NEBI_TaxID=46350;
CX

```

PN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=3H;
 RX MEDLINE=96266430; PubMed=8661429;
 RA Muramatsu S., Mizukami H., Young N.S., Brown K.E.;
 RT "Nucleotide sequencing and generation of an infectious clone of adenovirus 3.";
 RL Virology 221:208-217(1996).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=3H;
 RA Muramatsu S., Brown K.E.;
 RL Submitted (FEB-1996) to the EMBL/GenBank/DBJ databases.
 DR EMBL; U48704; AAC55049.1; -
 DR InterPro; IPR001403; Parvo coat.
 DR Pfam; PF00740; Parvo coat; 1.
 SQ SSOURCE 736 AA; 51660 MW; AF11EF475C67A10 CRC64;

Query Match	6.1%;	Score 110;	DB 12;	Length 736;
Best Local Similarity	22.4%;	Pred. No. 6.5;		
Matches 77;	Conservative 49;	Mismatches 144;	Indels 74;	Gaps 22

QY 18 DYRPSAGRAETVANNYIRKKQOVYSHRDKQOMTEBQREWLSYGV--GVWVNSGQ 74
Db 4 DGYLPDW--LEDNLSGIREW--WALKGPQPKANQHQDNRRGLVLPGYKYLPGNGS 58

Qy 75-----YPNRLAFASFEDR-FKNEIXNG-RP-RSGETRAEFEGVAKSFDEEGFQA 128
 Db 59 LDKGEFVNADAAALEHDKAVDQQLKAGDNPLYLKNHADAEFQEKLFQ---DTSGGNLG 115

```

127 REVASVNNRALE--NAHDESEYLDNKKELANGDALKRMDARSEFYSAKRNIESFKEKN 184
116 RAVFQAKKRILEPLGLVEEAAKTAPGKK-----GAVDQSQPEPPDS 150

```

QY 185 G-----NHDEPKKAVISKHFWSGQDRSSADKKKIGDEAFKAP-GIGLVMSKDKN 23
 157 SIVGKSGKQPRKRL-----NFGQTGDSSEVPDPPLGEPPA---APTSGSNTWASGGG 208

240 LFNSTSTG-----ESTVNE DIGNFGAL EULADNIV-----TNGNF- IMAFNOSJGAM 200

Db 209 APMADNNAGADGVSSNGNHCDSQMLGDRVITTSFRT WALPYNNHLYKQISSQSGAS 267
 QY 289 HYESKFRNMSE--GYSPDRGAYVITFIPKSWNTAPDXKQGW 330
 Db 268 N--DNHYFGISTPMGTFDPNR--FHCHFSRDMQRL--INNWN 304

RESULT 12
 ID 007351 PRELIMINARY; PRT; 919 AA.
 AC 007351.
 DT 01-JUL-1997 (TrEMBLrel. 04, Created)
 DT 01-JUL-1997 (TrEMBLrel. 04, Last sequence update)
 DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
 DE Protease precursor.
 GN PRY.
 OS Vibrio cholerae.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Vibrionales;
 OC Vibrionaceae; Vibrio.
 OX NCBI_TaxID=666;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=017;
 RX MEDLINE=89013889; PubMed=3050359;
 RA Alm R.A., Stroeber U.H., Manning P.A.;
 RT "Extracellular proteins of *Vibrio cholerae*: Nucleotide sequence of the
 RT structural gene (hlyA) for the haemolysin of the haemolytic EL Tor
 RT strain 017 and characterization of the hlyA mutation in the non-
 RT haemolytic classical strain 569B."
 RL Mol. Microbiol. 2:481-488(1988).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=017;
 RA Manning P.A.;
 RL Submitted (JUN-1997) to the EMBL/GenBank/DBJ databases.
 DR EMBL; Y00557; CAA6636.1; .
 DR MEROPS; M06.002; .
 DR InterPro; IPR000601; PKD domain.
 DR InterPro; IPR006025; Zn_MTPcdse.
 DR Pfam; PF00801; PKD; 2.
 DR SMART; SM00089; PKD; 2.
 DR PROSITE; PSS0093; PKD; 1.
 DR PROSITE; PSS0142; ZINC_PROTEASE; 1.
 DR PROSITE; PSS0142; ZINC_PROTEASE; 1.
 FT SIGNAL.
 FT SIGNAL.
 SQ SEQUENCE 919 AA; 101996 MW; 6A80774801FEBD8D CRC64;

Query Match 6.0%; Score 109; DB 2; Length 919;

Best Local Similarity 21.8%; Pred. No. 10; Mismatches 111; Indels 78; Gaps 16;

Matches 67; Conservative 52; Mismatches 111; Indels 78; Gaps 16;

QY 42 YSHR-DGRKQMTTEQREWLISYGVGTWNSQYPTNRL-APASFDD-RFKNELKNGR 98
 Db 446 YSNRGDLKRMNR-----LTI PASQATLFRKWPQIEKDYVAVILNGK 493
 QY 99 PRSGETAEFE---GRAKESFDEKGFQZAR-EVASVNDALNAD-----ES 144
 Db 494 PIAGNITMDPDKSLVPAISGDSGWDAGFDISAWAGQVELADYILDGLAMG 553
 QY 145 AYLDNLKKELANGNDAIRNEDARSPYSALRNTPSR-ERNGNHDSRKAVYSKHP 202
 Db 554 LYVDLRLLEVNGQTLIDNASTSF-----AQGFTKNGGHE-----ANHY 596
 QY 203 WSGQDRSSS-----ADKKKYGDPPAFRPAFTGLVMSRDNNIPRSPISGEGFVNFY 256
 Db 597 YLLQWRSHNDVDCGLANIKRFQGLMSPFGLVWYVDSEYADNWTGK--HFGG-----647
 QY 257 GMPGAQTADADKTVTTHGNHAPNGSLGAMHYESKPRMNSBGYSFPGAGVITFIP 316
 Db 648 GMLG-RVADADQNALVWSKTG-----EVAQTRFQVADATFSLFDQAP--LKLTV 692
 QY 317 KSWNTAPD 324

Db 693 ADGNTLED 700

RESULT 13

ID 096C72 PRELIMINARY; PRT; 401 AA.

AC 096C72; 01-DEC-2001 (TrEMBLrel. 19, Created)

DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)

DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)

DE Hypothetical protein (Fragment)

OS Homo sapiens (Human)

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.

OX NCBI_TaxID=9606;

RN [1]

RP SEQUENCE FROM N.A.

RC TISSUE=Muscle;

RA Strausberg R.;

RL Submitted (SEP-2001) to the EMBL/GenBank/DBJ databases.

DR EMBL; BC014625; AAH14625.1; .

KW Hypothetical protein.

FT NON_TER

SQ SEQUENCE 401 AA; 48615 MW; D9ADPDC029A0851D CRC64;

Query Match 6.0%; Score 108; DB 4; Length 401;

Best Local Similarity 21.5%; Pred. No. 4.1; Mismatches 68; Conservative 33; Mismatches 118; Indels 98; Gaps 12;

QY 39 QQYSHRDRKQMTTEQREWLISYGVGTWNSQYPTNRLAPASFDERFKNELKNGR 98
 Db 7 EERNHLEERKORKEER-----ITYYREKEEEOPRABEOMLKERERERARARA 57
 QY 99 PRSGETAEFEGRVAKESFDEKGFQZAR-EVASVNDALNAD-----ES 158
 Db 58 KRESELR-EYQERVKLEBEVERKKQRELEIEBERRR-----EZEERLD 102
 QY 159 DALRNEDAR-----SPFYSALRNTPSEK-RNGNHDPSRMKAVIYSKHPMS 204
 Db 103 SLSRDSRWGDRDSEGTWRKGPADSEVRGPEKEWRNRGSD----- 147
 QY 205 GQDRSSSADK---RKYGPD---AFRP---APGTVNDRDNINPRSPISGEGFVN 253
 Db 148 -EDRSRRDEERPRRLGDDEDEPSLFPDDRVPRKMDDRGPRGGEEDRSRGADD 206
 QY 254 FDYGMFQAQTB-----ADADKTVTTHGNHAPNGSLGAMHYESKPRMNSBGYSFDP 307
 Db 207 DRSNNTDDDPRIADDEDRGNWRADDDPRFRL-----DEBR 248
 QY 308 GAYVITFIPKSWNTAPD 324
 Db 249 G-----SWRTADE 256
 RESULT 14
 ID 09H0G5 PRELIMINARY; PRT; 558 AA.
 AC 09H0G5; 01-MAR-2001 (TrEMBLrel. 16, Created)
 DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
 DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
 DE Hypothetical protein.
 GN DKFZP434K1421.
 OS Homo sapiens (Human)
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Testis;
 RX MEDLINE=21154917; PubMed=11230166;
 RA Wiemann S., Weil B., Wellenreuther R., Gassenhuber J., Glasel S.,
 Ansoerge W., Boecher M., Bloecker H., Bauersachs S., Blum H.,

RA Lauber J., Duesterhoeft A., Beyer A., Koehrer K., Strack N.,
 RA Mewes H.W., Oltmannseder B., Obermaier B., Tampe J., Heubner D.,
 RA Mamuth R., Kohn B., Klein M., Poustka A.;
 RT "Towards a Catalog of Human Genes and Proteins: Sequencing and
 RT Analysis of 500 Novel Complete Proteins Coding Human cDNAs";
 RL Genome Res. 11:422-435 (2001).
 DR EMBL: AL136806; CAB6740.1; -.
 KW Hypothetical protein.
 SQ SEQUENCE 558 AA; 66390 MW; 99B7BDCFD06F98D CRC64;

Query Match 6.0%; Score 108; DB 4; Length 558;
 Best Local Similarity 23.8%; Pred. No. 6.3;
 Matches 59; Conservative 31; Mismatches 86; Indels 72; Gaps 11;

QY 36 RKWQVYSHDGRKQMTCEQEWLSYGVVWNSGQYPTNRLAFASPDREKFKELK 95
 DB RDQENHYTDYDKENSHRHE-----ASHDSHWKHEG 361
 QY 96 NGRPRSGETPAFEGVAEESFDEKGFQAPRE--VASVNRPALENADH--SAYLDN 149
 DB 362 EDKPRARDQR-ERSDPRVWREKREKYSQREQEDRQNDQNRPSKGEKEKSKAKEH 420
 QY 150 LK--KELANGNDALRNEDAR-----SPFYSA-----LRNTPS 179
 DB 421 MKVAKERYENDKYRPREKREYGVGVSERNODRKSSPNSRAKDKFLDQERSNMRMAK 480
 QY 180 FKERNGNHDPSPMKAVIYSKHPWS--GQDRSSADKRYKDPDA--FRPAPGTGLVDM 234
 DB 481 DKERN--QEKPSNSESLGAKRLTEBQEKKEQGER---PPEAVSKFAKRNNEETVMS 534
 QY 235 SRDKNIPR 242
 DB 535 ARDYLAR 542

RESULT 15
 ID 056139 PRELIMINARY; PRT; 736 AA.
 AC 056139;
 DT 01-JUN-1998 (TREMblrel. 06, Created)
 DT 01-JUN-1998 (TREMblrel. 06, Last sequence update)
 DT 01-JUN-2001 (TREMblrel. 17, Last annotation update)
 DS Capsid Protein VP1.
 CS Adeno-associated virus 3B.
 CC Viruses; ssDNA viruses; Parvoviridae; Parvovirinae; Dependovirus.
 CX NCBI_TaxID=68742;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=98080418; PubMed=9420229;
 RA Rutledge E.A., Halbert C.L., Russell D.W.;
 RT "Infectious clones and vectors derived from adeno-associated virus
 RT (AAV) serotypes other than AAV type 2.";
 RL J. Virol. 72:309-319 (1998).
 RN [2]
 RP SEQUENCE FROM N.A.
 RA Rutledge E.A., Russell D.W.;
 RT Submitted (SEP-1997) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AF028705; AAB95452.1; -.
 DR InterPro: IPR001403; Parvo_coat.
 DR Pfam: PF00740; Parvo_coat; 1.
 SQ SEQUENCE 736 AA; 81906 MW; D52331AD5FD70F CRC64;

Query Match 5.9%; Score 107; DB 12; Length 736;
 Best Local Similarity 22.5%; Pred. No. 11;
 Matches 76; Conservative 47; Mismatches 153; Indels 62; Gaps 21;

QY 18 DPYRSTGRATVNNYIRKQOVYSHDGRKQMTCEQEWLSYGVV--GYVWNSGQ- 74
 DB 4 DGYLPDW--LEDNISGIREW---WALKPVGPQKXQOCHODNRGLVLPQKYLGPNG 58
 QY 75 ----YPTNRLAFASPDREK-FKDELKNG-RP--RSGETPAFEGVAEESFDEKGFQRA 126
 DB 59 LDKPEPVNADALALSHDKAIDQQLAGNDYILKINHADAEPQERLQE---DTSGGNLG 115

QY 127 REVASVNRALNADHESAYLDNIKELANGNDALRNEDARSPFYSALENTPSEKERNG 186
 DB 116 RAVPQAKKRILBPL-----GLVEAAKTAQKXRPVDSPP-----QEPDSSSGVGKS 162
 QY 187 NHDSSRMKAVIYSKHPWSGQDRSSADKRYKDPDAFRAP-GTGLVDMERDRIIPASPT 245
 DB 163 GKQPARREL----NFGQTGDSBSVPDPQPLGEPPA---APTSLGNTMAAGGAPVADN 214
 QY 246 SPG-----EGFVNPDYGMFGAQTEADADKTVM---THGNH-YHAPGSLGAMHYESK 294
 DB 215 NEGADGVGNSGQWHDSCWLDGRVITTTSTRT-WALPTYNHLYQISSGGSASN--DNH 271
 QY 295 FRMWSE--GYSDPDGAYVITFIPKSNNTAPDKVKQGW 330
 DB 272 YFGYSTPWGYFDENR--PHCHSPRDMQRL---INNWN 304

Search completed: September 22, 2003, 12:45:04
 Job time : 44 secs